

A chance for change in France

A budget crisis is crippling French science, despite the best efforts of the research minister. But the time is right for a radical and necessary reorganization of research.

Science has been making headlines in France's newspapers, but not for the reasons that some might like. Laboratories are suffering from brutal and indiscriminate cuts by the centre-right government, which has lopped as much as a third off some labs' budgets. Whatever the government's economic predicament, these cuts and the resulting job losses are sending the wrong signal to the young minds that French science badly needs to attract.

No wonder scientists are livid. Electoral pledges by President Jacques Chirac last year had left them expecting research to be a national priority, with annual spending being hiked from 2.2% of gross national product to 3% before 2010, in line with goals agreed by European Union heads of states (see page 6). Claudie Haigneré, the science minister, is sympathetic to the scientists' plight, and recently won a reprieve from a further round of cuts. But the best she can probably hope for is that a vague promise by Prime Minister Jean-Pierre Raffarin will lead to a better budget next year. What irks many is that the nation's scientific and technological fabric appears to be being sacrificed to pay for populist electoral promises such as tax cuts and a dubious spending spree on the police.

The critics have a point. The science-budget crisis is a symptom of an increasing political disengagement over recent decades. Now the ship is adrift without a captain. Substantial progress demands a renewed and sustained political commitment, and a new contract between science and government built on mutual confidence. Scientists themselves must shake up the bloated and bureaucratic French science system, so that it provides clearer goals, flexibility and cost-effectiveness, and responds to the aspirations of young scientists.

This will require strong and constructive contributions from all stakeholders, including the trade unions and a new and positively

inclined generation of research leaders. Haigneré is contemplating a radical reform, transforming the country's large research organizations that run their own laboratories, so that they play a greater role as research councils, funding autonomous laboratories through competitive peer review. None of her predecessors succeeded in pulling this off, often because they tried to transpose and impose superficially appealing Anglo-Saxon models without taking into account the cultural realities of the French system.

Adopting what works well elsewhere would liberate French scientists from stifling centralization. Encouragingly, Haigneré seems to appreciate that change will require careful thought, as France's weak university system is in no fit state to manage the country's laboratories — progress may require a hybrid system involving the research organizations. She is in an unenviable position, but seems to have a sincere desire to listen, and to seek constructive ways forward; scientists and their trade unions should work with her, not against her.

But researchers can only do so much. Meeting the challenges of twenty-first-century science in France requires political leadership that is currently lacking. No French politician since has matched the powerful vision of the role of science and technology pioneered by Pierre Mendès France, who led a Socialist government in the early 1950s, and General Charles de Gaulle, who returned to power in 1958. Despite the competing demands of the need to build the social-security system, and fighting wars in Indochina and Algeria, France's governments of the time turned that vision into an unprecedented expansion of science, taking France and Europe to the leading edge of molecular biology, space and nuclear technology. France, under Chirac, now has a similar opportunity — and for the sake of its and Europe's international status in science, France must take it. ■

Rice institute needs strong support

Despite rumours to the contrary, the role of the International Rice Research Institute is as important as ever.

Rice has been a great success story. Since the Green Revolution in the 1960s, rice production has more than doubled, thanks to a 2.5-fold increase in productivity per hectare. Vietnam, which once struggled to meet its needs, is now an exporter. There is now more than enough rice to go around, but the eastern regions of India, suffering floods and soil alkalinity, struggle to meet their own needs despite the abundance of rice produced in the well-irrigated Punjab region. Telling people to redistribute rice won't help much. Local growers need to be able to look after themselves — for them, research into productivity continues to play an essential role.

The International Rice Research Institute (IRRI) near Manila in the Philippines is adapting to these circumstances. Rather than offering a finished product that can be used in any region, the institute is developing materials, methods and training so that breeders can take advantage of genomic tools to meet their local needs, whether the soil is acidic or alkaline, and whether the regions are prone to floods, drought or both (see *Nature* **422**, 796–798; 2003).

The problems facing farmers will only get worse. Rice demand is

expected to increase at 1% per year over the next 25 years. As sea levels rise and soil becomes salty and eroded, even greater productivity will be necessary. Nearly half of the diverted water in Asia goes to rice irrigation, but water, too, is becoming scarcer and tainted.

Researchers hope to tap the secrets of the rice genome to meet these challenges — a good bet, considering the unexplored biodiversity in the rice germ stocks. But there are significant obstacles to bringing genomic science to bear on farmers' practices. IRRI, whose rice lines have been bred into over a third of the new lines produced worldwide since the 1960s, is well positioned to take up that challenge.

But the institute is facing huge cutbacks (see *Nature* **416**, 777; 2002). In the three years from 2001 to 2003, IRRI's annual core funding dropped by 26%, and similar cuts are expected in the future.

It is essential that support for IRRI be mobilized. Researchers there, where research that spurred the Green Revolution was carried out, sometimes hear their success in producing abundant, high-yielding rice as a justification for cutting their budget, as if to say "your job is over". But the institute's job is not over — it has just begun. ■





Powered up
Neutron source set
to make use of
weapons-grade fuel
p5



Raising Arizona
Bioterror adviser
to head biodesign
institute
p6



Sound of silence
Quiet ship will
collect data by
creeping up on fish
p7



Hot topic
Thermal cameras
seek fevered brows
of SARS sufferers
p8

Apartment complex holds clues to pandemic potential of SARS

**David Cyranoski, Tokyo
and Alison Abbott, Munich**

A 33-floor apartment block in the bustling Kowloon district of Hong Kong could provide clues to the true risk posed by severe acute respiratory syndrome (SARS), epidemiologists say.

Scientists around the world are racing to gather data that will allow epidemiologists to predict whether SARS is likely to become a full-blown pandemic. Their search has led them to investigate exactly what happened when the disease surged through the Hong Kong residential complex in March.

The study, which is being led by Wilina Lim, head of virology at the Hong Kong Department of Health, is the world's only large-scale investigation of the extent to which people without symptoms are carrying SARS — a critical question for those seeking to predict the epidemic's likely reach.

"The key issue is whether symptomless carriers can infect others, and for how long they can do so," says Albert Osterhaus, head of the Netherlands National Influenza Centre in Rotterdam. "If carriers are symptomless for extended periods, a pandemic is quite possible. If not, the epidemic is likely to be controlled."

As of 28 April, 131 cases of SARS had been confirmed among residents of Block E of the heavily hit Amoy Gardens apartment complex in Kowloon. And 241 symptomless



Block E of the Amoy Gardens residential complex is the focus of a major Hong Kong investigation.

residents were quarantined on 1 April for 10 days following the local outbreak.

Just over 100 of the quarantined residents are taking part in Lim's study, the results of which, she says, should be known by mid-May. The researchers are trying to take throat samples from the participants every other

day to follow the course of the disease, should it emerge. The samples will be tested for SARS using a version of the polymerase chain reaction (PCR) to amplify and detect genetic material from the coronavirus that is thought to cause the illness. But Lim admits that since the end of the quarantine, sampling has been

Biologists seek to head off future sources of infection

Researchers at the Beijing Genomics Institute of the Chinese Academy of Sciences have mounted a wide-ranging search of domestic and wild animals, in a hunt for coronaviruses similar to that which is thought to cause severe acute respiratory syndrome (SARS). The work is part of a broader international effort that is attempting to seal off potential reservoirs of future infection.

"Dozens of species have been tested," says the institute's director, Henry Yang, although he declines to reveal which ones, and whether they have been found to be free of coronavirus. "If we

said we haven't found anything yet in, say, chickens, people would rule out the chicken as the reservoir," he says. "But that would be jumping the gun."

The closest known relatives of the human SARS virus are coronaviruses from cows and birds, species that Yang says his group is treating as priorities. Coronaviruses have also been found in pigs, mice and cats. But the relationship between these animal viruses and the human SARS virus is relatively distant — so Yang also intends to look in other species, such as bats and monkeys, to see whether

a closer relative can be found.

Meanwhile, other researchers are looking at the possibility that SARS is transmitted in blood — which would make blood banks potential reservoirs. As yet, they do not know whether the virus is present in the blood of symptomless carriers, although most consider this unlikely by extrapolation to other viral infections. Agencies such as the US Food and Drug Administration and the Frankfurt Blood Bank are looking at screening strategies to identify the virus, should it turn up in blood.

Alison Abbott and David Cyranoski

less consistent as the subjects do not always return to the hospital as requested.

Blood samples were also taken from each of the subjects at the beginning of the quarantine period, and they will be taken again this week. These will be tested for the presence of antibodies against the coronavirus, which will show whether the subjects have been exposed to the virus and have mounted a successful defence. Lim is not prepared to confirm how many of the study's subjects have now developed symptoms of the disease.

Data from these samples, as well as from patients suffering from SARS, will offer clues to some of the key biological questions, says Lim, including when the virus starts to show up in bodily fluids, where in the body it comes from, and whether there are symptomless carriers. "At the moment we don't know enough about the route of transmission of the virus," Lim says. "There is just not enough material yet for investigation."

The study has also analysed hundreds of samples taken from surfaces in the Amoy Gardens apartments, such as toilet seats and door handles, some of which have tested positive for SARS in PCR tests.

In the next few weeks, Lim's group will also begin a study on people from outside Amoy Gardens who have had contact with the disease but have yet to display symptoms.

Other researchers involved in the global network established by the World Health Organization (WHO) in response to the SARS outbreak confirm that it will be several weeks, or even months, before they will



Covered up: students in Kowloon protect themselves following the area's outbreak of SARS.

have enough data to make even the broadest of predictions about the disease's spread.

Samples from patients from countries where SARS infection has been significant, including Hong Kong, Vietnam and Singapore, are also being tested regularly to determine how best to track the virus during the course of the disease. Results of PCR analysis show that stool samples test positive for viral particles for up to 30 days, by which time symptoms have often disappeared. The particles seem to be infective for up to ten days — once isolated they can infect normal cells in culture — but it has not yet been established whether they stay infective after that.

Tracking populations of sick patients and healthy people who have been in contact with them is a powerful method for trying to understand the progress of SARS, but many details can only be investigated in animal models. A monkey model based on

macaques, which has just been developed at Erasmus University in Rotterdam, the Netherlands, is now also being used at the US Centers for Disease Control and Prevention in an attempt to understand factors such as how quickly symptoms appear after infection.

But with primates so expensive and awkward to work with, labs in the WHO network are trying to develop models for SARS in more convenient animals such as mice. "We'll need several months to get a suitable small animal model," says Christian Drosten, of the Bernhard Nocht Institute for Tropical Medicine in Hamburg, Germany.

The WHO, meanwhile, has already issued strict travel guidelines, assuming the worst. "We don't have the luxury of seeing what happens and then acting," says David Heymann, the WHO's executive director for communicable diseases.

♦ www.who.int/csr/sars/travel

Critics slam treatment for SARS as ineffective and perhaps dangerous

Health authorities in Hong Kong are coming under intense criticism over their use of an antiviral drug combined with anti-inflammatory steroids to treat patients with severe acute respiratory syndrome (SARS).

Critics of the treatment say that there are few indications that the combination of steroids and the antiviral drug ribavirin is effective. And autopsies revealing damage to organs other than the lungs have raised the question of

whether the drugs are partly to blame.

"There has been concern in Hong Kong about the efficacy of the treatment, and doctors there are calling for a re-examination of the protocol amid concerns that the prolonged use of high-dose ribavirin may be causing or exacerbating multi-organ failure," says Gurinder Shahi, a physician and chairman of the Singapore-based biotechnology company BioEnterprise Asia.

Many in Hong Kong say that the necessary clinical tests have not been done, and that it is therefore too early to judge. "There is no clear evidence either way right now," says Ka-fai To, a pathologist at the Chinese University of Hong Kong whose autopsies of SARS victims found evidence of damage to the lymph nodes and spleen. "These findings could be side effects of the drug, but they could also be the effect of the virus or even an immune response to the virus," she says.

In Hong Kong, which has the largest number of SARS cases and deaths outside mainland China, all patients receive the combination treatment, says John Tam, director of virology at the Prince of Wales Hospital in Hong Kong. The combination is also being used in China's

Guangdong Province, Singapore and Toronto.

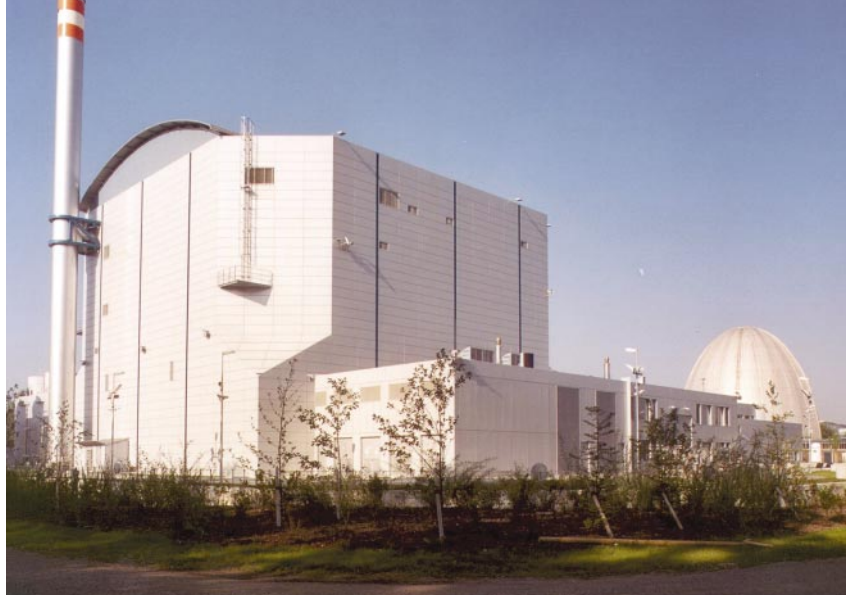
Although ribavirin is known to be toxic in some patients, it has been prescribed because of its effectiveness against respiratory syncytial virus. This virus is similar to the metapneumovirus that was first isolated in Hong Kong and was originally thought to be the cause of SARS, says Masato Tashiro of the National Institute of Infectious Diseases in Tokyo.

But the drug's efficacy against SARS, which is now thought to be caused by a new strain of coronavirus, is unknown. "Increasing numbers of patients show no improvement," says Tam. *In vitro* tests are also negative, according to Tashiro.

But in the absence of any obvious alternative, health authorities in Hong Kong are sticking with the therapy. In a speech in Kuala Lumpur on 26 April, the health secretary Carrie Yau Tsang Ka-lai said that 80% of patients seem to respond favourably to the treatment.

But US officials have rejected ribavirin as a treatment for SARS. At a press briefing on 22 April, Julie Gerberding, director of the Centers for Disease Control and Prevention, said: "There doesn't seem to be any activity of ribavirin against this particular coronavirus." David Cyranoski





Fuelling controversy: the FRM-II research reactor is likely to use highly enriched uranium until 2010.

Neutron source powers ahead with weapons-grade uranium

Quirin Schiermeier, Munich

The German government has given the go-ahead for the operation of a research reactor that will be fuelled by highly enriched uranium (HEU).

The decision, announced on 15 April, ends years of argument over whether scientists should be allowed to use a reactor whose fuel could be readily diverted to make nuclear weapons. But the government is demanding that by 2010 the FRM-II research reactor at Garching, near Munich, should be converted to use low-enriched uranium.

Despite continuing opposition from non-proliferation experts and nuclear watchdog groups, the US\$400-million facility is now likely to become fully operational this year. The facility was ready in 2001, but has been mothballed pending the decision.

The FRM-II is the world's first large research reactor to open in the past 25 years that burns HEU, a weapons-grade material containing more than 90% of the isotope uranium-235 (see *Nature* **414**, 5; 2001).

The 20-megawatt facility will give physicists, chemists, biologists and engineers access to an intense flux of neutrons for investigating materials and molecules. It incorporates 14 beam lines and 5 other radiation facilities for such experiments, including a unique source of 'ultracold' neutrons and a powerful positron probe. Ultracold neutrons move very slowly and are useful both for the study of neutrons themselves and for other applications, such as optics. The positron probe will be used to investigate atomic-scale structure and defects in solids.

"We'll finally be able to show what

fantastic opportunities neutrons have to offer, and we can begin to train the next generation of neutron researchers," says Winfried Petry, scientific director of the FRM-II.

But not everyone in the neutron community is as thrilled as Petry. Some contend that the risks surrounding the use of HEU are too high a price to pay for the expected scientific gain.

"Basically, this is the technology of the 1960s. There is absolutely no scientific requirement any longer for using HEU in a research reactor," says Franz Fujara, a physicist at the Technical University of Darmstadt who has been involved in neutron research for more than 25 years. Fujara says he will not apply to use the FRM-II as long as it is run on weapons-grade fuel.

Converting the reactor to use low-enriched uranium fuel will reduce its neutron flow density, according to a technical advisory committee to the government, but will not significantly limit its scientific usefulness (see *Nature* **400**, 7; 1999).

But the German government has not yet specified just how low the uranium enrichment must be after the conversion. Under international agreements, only uranium containing less than 20% uranium-235 is considered non-weapons-grade, low-enriched uranium (LEU).

"We are working together with industry on the development of a new high-density LEU fuel," says Petry. "But there is no guarantee that a fuel element that suits the FRM-II's design will be available by 2010."

♦ www.frm2.tu-muenchen.de/index_en.html

Human fatality adds fresh impetus to fight against bird flu

Alison Abbott, Munich

The death on 17 April of a veterinarian who contracted flu from chickens in the Netherlands has heightened concerns about the threat posed by the disease.

Since the outbreak of the illness in late February (see *Nature* **422**, 247; 2003), 18 million chickens have been culled in the Netherlands — including six million killed last week close to the German border. But the flu has now spread to birds in Belgium, prompting the slaughter there of hundreds of thousands of chickens.

Transmission of the disease to humans remains a major concern. Ron Fouchier, a virologist at Erasmus University in Rotterdam who monitors human cases of the avian virus, says that he "remains very worried" about the outbreak.

Since 28 February, 83 people have become infected with the virus, known as H7N7. Most have suffered only conjunctivitis, but eight have developed mild flu-like symptoms. The vet, who was from Den Bosch, died from pneumonia.

"The numbers are, of course, small — but the death rate among those with flu-like symptoms is similar to that for the Hong Kong outbreak in 1997," says Fouchier. In Hong Kong, 18 people suffered flu-like symptoms and six died from a bird-flu virus identified as H5N1.

Neither H7N7 nor H5N1 spreads rapidly between humans, so a pandemic is thought to be unlikely. "But we don't have things under control," Fouchier says.

The number of new cases in humans has slowed in the Netherlands since 10 March, when at-risk workers began taking the antiviral drug Tamiflu. Belgium has also introduced measures to protect people working with chickens.

Marion Koopmans, a spokeswoman for the RIVM, the Netherlands' public-health laboratory, warns that other countries should be on alert, particularly as the chicken cull involved several thousand foreign labourers.



Disease burden: 18 million Dutch chickens have been culled in the fight against bird flu.

Commission lays foundations for rise in research spending

Declan Butler, Paris

The European Commission says that it will allow "small and temporary" budget deficits to be run up by member states that are trying to boost spending on research.

The move is seen as lending support to a plan by the European Union (EU) to boost its total research and development (R&D) spending from 1.9% of economic output to 3% by 2010. Normally, member states that run budget deficits above target levels face huge fines from the commission.

Increased public spending on research is needed if the EU is to encourage more private investment; it is relying on industry to come up with two-thirds of the proposed R&D increase.

But even as the commission announced its offer, a survey by the European Round Table of Industrialists, a forum of 45 leaders of Europe's largest research-based companies, found that most of its members plan to increase research activity in the United States or Asia, rather than in Europe.

The commission's offer is part of an action plan released this week aimed at meeting the EU's R&D target. The goal is seen as vital by Europe's politicians if the continent is to compete with the United States and Japan, which both spend about 3% of their economic output on R&D.

The commission's plan calls for EU member states to agree measures to provide tax breaks for industry, to strengthen links between industrial and publicly funded research, to redirect public spending towards research and innovation, and to make careers in research more attractive. It also suggests that member states convert fragmented national research initiatives into Europe-wide programmes in critical technologies such as aerospace and nanotechnology.

But there is still a long way to go to meet the R&D goal. Spain currently spends only 0.97% of its economic output on R&D, lagging behind Italy (1.04%) and Britain (1.84%). France (2.13%) and Germany (2.52%) are closer to the target, but only Finland (3.37%) and Sweden (3.78%) surpass it.

The commission itself has little direct clout to enforce change, as member states set taxes and provide most public research funds. Officials say that it will rely instead on encouraging best practice in innovation and research policy, and on monitoring closely the performance of member states.

French researchers demand radical overhaul of funding

Declan Butler, Paris

Major reforms to decentralize France's research system were backed by most speakers at a high-level meeting on 28 April.

About 50 leading scientists and officials, including science minister Claudie Haigneré, met in Paris to discuss science policy. The meeting took place against a backdrop of protests by researchers against government budget cuts that are forcing the government-run laboratories that form the backbone of French research to deal with reductions of up to one-third in their operating budgets.

While attacking the cuts, speakers said that the time was ripe for a radical change in how French science is run. Philippe Kourilsky, director general of the Pasteur Institute in Paris said that the malaise in the system is palpable.

Christian Bréchet, director general of INSERM, the national biomedical agency, was even blunter: "It's obvious that our model doesn't work," he said.

Many speakers expressed admiration for what is seen in France as the 'Anglo-Saxon model' of competitive grants based on peer review, as opposed to the French system in which most of the available money directly funds the laboratories of the main research organizations. Haigneré indicated that although such a model is not "directly transposable" to France, the government would like to introduce some elements of the system, turning the existing research



Claudie Haigneré: considering science reforms.

organizations into granting agencies.

Greater collaboration within the European Union was said by several speakers to be the best hope for France and its neighbours to strengthen their national research efforts, although concerns were expressed about the quality of existing European Commission research programmes.

Faced with a revolt by researchers, Prime Minister Jean-Pierre Raffarin recently promised that President Jacques Chirac's pledge to increase spending on research from 2.13% to 3% of economic output by 2010 would be met, with government spending on track to hit the target by 2007.

Arizona institute names leader

Rex Dalton, San Diego

Arizona State University (ASU) has hired a scientific heavyweight to head an ambitious, multidisciplinary research centre being built on its Tempe campus.

George Poste—former head of research at drug firm SmithKline Beecham and adviser to the British and US governments on bioterrorism—is set to be named this week as the new director of the Arizona Biodesign Institute (AzBio). Research projects at the institute are expected to include nanotechnology, materials science, robotics, information technology and drug-production methods.

Trained as a virologist and veterinarian in Britain,



George Poste: five years to "make things happen".

Poste lives in Arizona. He is a member of the US government's Defense Science Board and chairman of its biodefence task force.

Fifty-eight-year-old Poste says that he has given the ASU "a minimum five-year commitment to make something special happen".

His appointment was due to be announced at a ceremonial groundbreaking on AzBio's first research building. The \$70-million structure is scheduled to be completed in late 2004, and the Arizona legislature is now debating a bill that would provide a further \$185 million for three more AzBio research buildings at the ASU.

"Poste is one of the leading scientists in the world, and there is no better person to build AzBio into a world-class centre," says Michael Crow, the ASU's president.

Charles Arntzen, a plant biologist who has been AzBio's interim director, will return to research, as was previously planned.

Columbia inquiry homes in on faulty foam

Tony Reichhardt, Washington

The official investigation into the loss of the space shuttle Columbia on 1 February is fast closing in on the accident scenario that seemed most likely in the first place. A piece of foam insulation, investigators believe, fell from the shuttle's external fuel tank shortly after launch, damaging the shuttle's left wing and causing it to overheat on re-entry.

The Columbia Accident Investigation Board, chaired by retired US Navy admiral Hal Gehman, compared notes with NASA's own investigators in a closed meeting on 24 April. The most likely scenario is that the foam damaged a T-shaped seal between two curved sections of insulation on the wing's leading edge. The seal then came loose while the craft was in orbit, creating a gap that allowed superheated gases to enter the wing when Columbia re-entered the atmosphere.

This scenario fits with evidence from the examination of recovered debris, data returned from the shuttle before it broke up, radar signals from material that floated away from Columbia in orbit, and data from a sensor package recovered in March, according to members of Gehman's board.

But there still are key gaps in the analysis. Whether falling foam could have caused the damage to the T-seal, even at high velocities, will be determined in part by impact tests to be carried out later this month. NASA and the Gehman board have been investigating areas on the fuel tanks where foam can come

loose, and the board may recommend that NASA revises its approach to applying the foam before shuttle flights can resume. Also under scrutiny are the reinforced carbon-carbon panels and T-seals, which may have weakened with age.

Finally, the board will consider NASA's decision-making processes, and whether shuttle managers should have recognized that past problems with falling foam were more serious than they had thought.

Meanwhile, a two-man crew arrived on the International Space Station on Monday for a six-month stay, on board a Russian Soyuz vehicle. Their supplies will be delivered by Russian Progress capsules.

Little research will be done on the station until the shuttle resumes deliveries of laboratory equipment, however. The next flight was to have brought up a second facility for research on human physiology, equipment for Earth observation, and a freezer for biological specimens. Subsequent flights would have boosted electrical power on the station by adding solar arrays.

But space-station crews can do research in the area of human adaptation to space, says Milburn Jessup, a professor of oncology at the Georgetown University Medical Center in Washington DC, and chair of the space-station scientific users' committee.

NASA is expected to take around a year to get the shuttle flying again. Michael Kostelnik, deputy head of the agency's shuttle and



Crumbling foam from the nose of Columbia's launcher (orange) may have damaged the craft.

space-station programmes, says a resumption of shuttle flights between next January and March is "hopefully possible". But scientists who are familiar with space-station planning say that flights are more likely to resume in summer 2004. ■

L. ALVAREZ/AP

Stealth ship sets sail for a quiet life fishing for data

John Moore, Cork

The world's quietest research vessel set sail last month on its first scientific mission. By using low-noise electric engines and mounting equipment on rubber, Ireland's £32-million (US\$35-million) *Celtic Explorer* should be able to approach fish more closely, improving the accuracy of fishery surveys.

Noise from ships is recognized as a major source of bias in such surveys, according to a

1995 report by the International Council for the Exploration of the Sea (ICES). Low-frequency sounds from vessels cause fish to stay 100–200 metres from some ships, and up to 400 metres from noisier ones.

Only two other vessels — the *Scotia* and the *Corystes*, both based in Britain — meet targets cited in the ICES report for reducing the sound from ships, and the *Celtic Explorer* outperforms them both. "Its performance is better than any fisheries research vessel measured to date," says the report's author, Ron Mitson of Acoustec in Lowestoft, UK, a company that advises ICES on underwater noise levels.

Most noise from ships is caused by pumps, engines and the propeller, or by protrusions on the hull that generate bubbles in the water. On the *Celtic Explorer*, all of the equipment is rubber-mounted onto the hull. The propeller mounting also reduces noise, by fixing the propeller in a set position, rather than allowing its angle to be adjusted as on some other vessels.

The ship's project manager, Michael Gillooly of the Marine Institute in Galway, Ireland, adds that the resulting low noise levels reduce interference in the sonar systems that are used to survey fish.

Celtic Explorer, which can accommodate 19 scientists, will conduct a seabed survey around Ireland until October, with the aim of producing maps for telecommunications applications, management of underwater hazards, and the protection of areas of scientific interest. In November, it will carry out a survey of fish such as cod and plaice that live on or near the bottom of the ocean.

Other countries outside Europe are also developing vessels that can fulfil the ICES guidelines. "We are currently in the middle of constructing the first of four new vessels designed and contracted to meet the ICES noise standards," says John Hotelling of the US National Marine Fisheries Service in Silver Spring, Maryland. "The first ship, *Oscar Dyson*, is expected to be launched in September and be in service by late 2004." ■



The strong, silent type: rubber mountings mean that the *Celtic Explorer* won't frighten fish.



Hot on the trail: heat-detecting machinery installed at New Tokyo International Airport.

Asian airports use thermal imaging to hunt down SARS

Tokyo A military-designed thermal-imaging system is being used to scan airline passengers entering and leaving Singapore. Similar systems are being introduced in Tokyo, Hong Kong and Beijing. Authorities hope that the technology will spot those with a high temperature brought on by severe acute respiratory syndrome (SARS).

In Singapore, for example, all passengers leaving the country since mid-April, and some entering, have been scanned with the infrared device. Those showing red on the monitor — meaning that their temperature is above 37.5 °C — are escorted to a nurses' station for a check with a thermometer. Anyone with a temperature above 38 °C is then taken to a hospital wing dedicated to treating SARS patients.

A high fever is one of the trademark signs of the disease, which had killed 21 people and infected 198 in Singapore by 26 April.

Critics of the imaging system have argued that it will pick up fevers associated with the common cold, whereas people who are infected with SARS but are not yet showing symptoms will pass through undetected.

Rocket-engine defect delays observatory launch

Washington The launch of the last of NASA's 'great observatories' — four telescopes covering different parts of the electromagnetic spectrum — has been delayed until August.

The Space Infrared Telescope Facility (SIRTF) had been due to launch on a Delta rocket from Cape Canaveral, Florida, on 18 April. But NASA announced on launch day

that engineers had discovered a defect in the material surrounding the exit nozzle on one of the rocket's nine engines.

The first of the agency's two Mars probes is scheduled for lift-off in early June, and officials have decided that SIRTF cannot be fixed before then. The telescope will now be launched no earlier than mid-August.

Royal Society to give prince nanotech news

London Prince Charles has asked the Royal Society to supply the names of scientists who could advise him on nanotechnology. The move comes after the ETC group, a Canadian environmental organization, published a report highlighting what it sees as the potential dangers of nanotechnology. The heir to the British throne has spoken out on other environmental issues, such as genetically modified crops.

The ETC report, published this January, claimed that nanotechnology — the manipulation of matter at the atomic level — may have unpredictable consequences and could be used to develop weapons of mass destruction. The report was greeted with scepticism by many researchers, but attracted media attention.

On 11 June the ETC group will run a seminar in Brussels, aimed at policy-makers, on the risk and promises of nanotechnology, and how it should be regulated.

Telescope loses space to Native American site

San Francisco A long-running wrangle over the construction of a telescope array near a site valued by Native Americans took another turn last week.

Construction of the Very Energetic Radiation Imaging Telescope Array System (VERITAS) appeared to have been given the

all-clear in January (see *Nature* 421, 466; 2003). The plans had generated opposition because the proposed site near Mount Hopkins in Arizona was close to a site used by Native Americans.

On 23 April Lucia Turner, deputy regional forester in Albuquerque, New Mexico, overturned the earlier decision on the grounds that VERITAS would be an eyesore. Officials working on the telescope, which would use seven 10–12-metre dishes to observe γ -ray bursts, say that they will appeal.

The Forest Service is now considering whether it should ask the telescope's backers to complete a potentially lengthy environmental impact survey.

Raelian cloners 'have two staff and no lab'

Washington Clonaid, the company that announced last December that it had cloned a human (see *Nature* 421, 3; 2003), seems to have no lab space and only two employees, according to court documents obtained by the *Boston Globe*.

Clonaid was founded by a religious sect known as the Raelians, who believe that the human race was cloned from aliens. The company's managing director, Raelian bishop Brigitte Boisselier, has said that so far a dozen clients have each paid \$25,000 or more for the group's cloning services.

The documents, which stem from a case in which a Florida lawyer tried to get a legal guardian appointed for the company's first purported clone, seem to show that the company is little more than a website, according to the newspaper.

Clonaid admits that its US operation is more or less just a trademarked name, but insists that research is being conducted at facilities overseas, which, it says, operate under different company names in order to preserve anonymity.

Anglers face the painful truth

London Does a rainbow trout on a hook feel pain? Yes, say researchers at the University of Edinburgh and the nearby Roslin Institute. The results could shatter the long-held belief that fish do not experience pain in the same way that higher vertebrates and humans do.

A team led by Roslin marine biologist Lynne Sneddon has found that rainbow trout have nociceptors — pain receptors on skin that report stimuli to the brain and cause reflex reactions when an animal's tissue is being damaged.

The group has also found that the animals react to painful stimuli in a way that cannot be explained as a reflex response, such as by producing rocking motions, a typical pain reaction in mammals.

"These are significant findings," says Michael Pietrock, a fisheries scientist at the Institute of



Inland Fisheries in Potsdam, Germany. The discovery is also likely to fuel arguments between animal-protection groups and anglers, he adds. Anti-angling campaigners have long believed that fish feel pain and cite this as the main argument against the sport.

The tiny toolkit

Can we probe the workings of cells without destroying them? Yes, says an influential and interdisciplinary group of US researchers — the answer lies in nanotechnology. Catherine Zandonella reports.

As the basic unit of life, the cell gets little respect. When we want to know what is happening inside it, we rip open its membrane and dump out its contents. And along with the cellular fluid and organelles, out of the cell leaks valuable information that is lost forever.

If we could go a little easier on cells, that information might become available to researchers. And according to the Nano-Systems Biology Alliance, a group of seven scientists on the US west coast, nanotechnology could provide the means to do so. They are working on tools that, although still in their infancy, could one day be capable of tracking life within the cell in real time.

The group's planned devices certainly sound impressive: arrays of nanowires that can detect thousands of proteins secreted by a cell are just one example. And beyond the alliance, other researchers are creating equally exciting inventions, including a nano-sized fibre-optic probe that can seek out target molecules, and gold particles that can be used to turn specific proteins on and off. Many of these projects may fail before they produce practical devices, but those that succeed will give biologists the chance to experiment on cells in a new way — as a system, rather than as a collection of organelles and individual processes.

To view a cell in this way — as a thriving beehive of interconnected activities such as cell signalling and the shuttling of nutrients

and metabolites — researchers need to think about their experiments in a new way, says Jim Heath, a nanotechnology researcher at the California Institute of Technology, Pasadena, and a member of the nanosystems alliance. He likens the process to a child playing a video game — instead of reading the instructions, they are likely just to pick up the joystick and go for it. When treating the cell as a system, experimenters are forced to do the same. “We can't approach this by reading the instruction manual starting at page one,” says Heath. “We need to jump in and play the game.”

Cell block

To begin with, researchers need the right controller. This could soon be available in the form of the nanolab, a device being developed by Heath and his colleagues that will combine several assays on a centimetre-square silicon chip. The chip resembles a miniature cell farm, with rows of cells, each nestling in its own well atop a tiny pore in the silicon. Fused with the cell membrane, the pore serves as a conduit between the inside of the cell and the outside world. Close to this conduit is a densely packed array of nanowires, metal rods just a few nanometres thick. Each nanowire is coated with a bimolecular probe, such as an antibody, that binds to a target protein. Proteins that diffuse through the membrane and bind to an antibody change the nanowire's

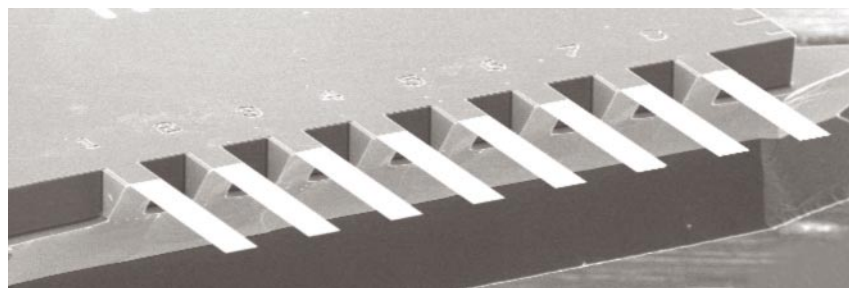
Light the way: molecular-sized probes that light up when they bind to a target ion are one of several promising new tools for cellular studies.

electrical conductance, and this can be measured by a detector connected to the array.

Several groups have already reported using carbon nanotubes and nanowires to detect specific DNA sequences and proteins. What is new about Heath's approach is that 1,000 nanowire detectors can be jammed into a few square micrometres — roughly the area taken up by a single cell. To create these arrays of nanowires, Heath and his colleagues developed a new technique, called superlattice nanowire pattern transfer, which can create individual semiconductor nanowires that are as little as 8 nm in diameter with the same distance between each wire¹. Previous methods could only produce 20-nm nanowires separated by gaps of 40 nm.

Potentially, each of Heath's nanowires could bear a different antibody or oligonucleotide, a short stretch of DNA that can be used to recognize specific RNA sequences. “On one chip, we will have 1,000 single-cell experiments,” says Heath. So far, he has built chip prototypes and devised methods for coating each wire with a different antibody. Over the next few months he plans to refine these techniques, and to begin to test the chip on the proteins secreted by cancer cells.

Some barriers remain to be overcome. The



A team at the California Institute of Technology is using tiny cantilevers to probe molecular bonds.

cells need to be surrounded by an artificial fluid that mimics their normal environment. But ions in this fluid can cover the sensor and obscure biochemical events that occur more than a few nanometres away. To get around this, researchers place cells in a fluid with a low density of ions, but this stresses the cells and makes the data less reliable. Advances in the surface characteristics of nanowires may, however, give rise to more sensitive biosensors, which could be used in more natural fluids. “We haven’t reached the fundamental limits of these detectors,” says Scott Manalis, a nanotechnology researcher at the Massachusetts Institute of Technology (MIT) in Cambridge. “There is definitely room for improvement.”

For the moment, the nanolab can measure only those proteins that are secreted by the cell. This problem can be solved — albeit in a way that destroys the cell. Stephen Quake, an alliance member and physicist at the California Institute of Technology, has added microfluidic channels — tiny liquid-filled canals — to the slab, which can be used to shunt cells to different areas. In one part of the chip, chemicals that disrupt the cell membrane will be used to release the contents onto another array of nanowires, coated with DNA probes, to provide a snapshot of gene expression at that moment.

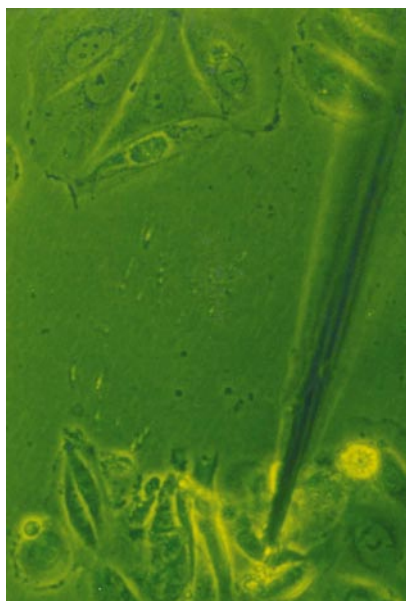
Another physicist in the alliance, Michael Roukes, also of the California Institute of Technology, is working with Heath and Quake on a section of the nanolab chip that will measure the binding forces between individual molecules. These forces can already be assessed using atomic-force microscopes (AFM), which generate atomic-scale images of surfaces by measuring the way in which a surface attracts and repels the end of a tiny metal cantilever. The force between, say, a drug and a receptor molecule can be measured by attaching the drug to a glass slide and the receptor to the microscope’s cantilever.

Roukes plans to implement something similar on the nanolab. Target receptors will be tethered to the chip, and the drug anchored on a cantilever. Roukes aims to determine the strength of the bond by measuring the change in springiness of the cantilever when the bond between the two molecules is severed. The change will be measured by a piezoelectric crystal — a device that converts mechanical pressure into electrical signals. Other groups

have used cantilevers to detect biomolecules, but Roukes’ devices are far smaller, and hence can generate a more detailed view of the forces between the two molecules.

To ensure that physicists such as Roukes and Heath produce genuinely useful tools, the alliance also includes three biologists. Charles Sawyers, a cancer researcher at the University of California, Los Angeles, already has plans for the nanolab. He wants to find out how chronic myeloid leukaemia cells become resistant to Gleevec, a relatively new drug that blocks an enzyme involved in the proliferation of cancerous cells. Sawyers suggests that the nanolab could be used to work out which genes are switched on when leukaemia cells are treated with the drug. “You could ask what is the first thing the cell does when it sees Gleevec,” Sawyer says. He hopes that this approach could reveal how the cells develop resistance.

Another alliance member is Alan Aderem, an immunologist and cell biologist at the Institute for Systems Biology in Seattle, Washington. He is interested in immune cells, such as macrophages, that display proteins called toll-like receptors on their surfaces — these recognize various pathogens such as bacteria



Good point: an optical nanofibre can probe a cell’s molecular contents without killing it.

and viruses. The nanolab would allow Aderem to trap a single macrophage and study which of its genes are being transcribed in response to the pathogens that it encountered. “If we could understand the information embedded in macrophages, we could use them as sentinels to see what is going on in the body,” he says.

Researchers outside the alliance are also working on new nanotechnologies. To take a closer look at a proteins, for example, it would be nice to enter the cell and poke around with a flashlight. Tuan Vo-Dinh, a chemist at Oak Ridge National Laboratory in Tennessee, is trying to do just that². His device is a fibre-optic probe, about 40 nm across at the tip, capped with an antibody that binds to a target molecule. Vo-Dinh makes the nanofibre by drawing out the tip of a normal optical fibre to a very fine point. The fibre is then coated with silver to stop light escaping from the sides.

Top tip

When Vo-Dinh turns on the device, light travels down the fibre-optic probe. Because the tip of the probe is much smaller than the wavelength of light, photons cannot go all the way to the tip. Instead, the photons go as far as they can and then create evanescent fields — waves of light that decay rapidly and hence only excite molecules at the tip of the probe. To the antibody at the tip, Vo-Dinh attaches a fluorescently labelled target molecule, and then sticks the probe into the cell. There, the cell’s copy of the molecule displaces the fluorescent version, causing the fluorescent signal to diminish and allowing Vo-Dinh to ‘view’ individual molecules.

This method has an advantage over simple fluorescent dyes, which can be used to label and track molecules, but eventually kill cells. “The nice thing is that when you pull out the fibre, the cells still survive,” says Vo-Dinh. He is now using the nanofibre to investigate apoptosis, or programmed cell death.

Nanotechnology also offers researchers the chance to detect rare events or molecules that are present only at low concentrations. Quantum dots — nano-sized crystals of semiconductor metals — emit intense light at specific frequencies and can, for example, be used to label a range of biomolecules³. Other similar technologies are under development. A team at the University of Michigan in Ann Arbor has developed PEBBLES, or Probes Encapsulated By Biologically Localized Embedding, which can be used to track the movement of zinc around cells and could aid studies of neurological diseases.

PEBBLES consist of fluorescent dyes trapped inside larger cage molecules, 20–200 nm in diameter. The dyes glow when they bond with zinc and, through a microscope, the PEBBLES can be seen winking on and off like fireflies, allowing researchers to track how and where zinc is stored and released in the cell. Abnormal zinc regulation is a feature of Alzheimer’s disease, but current tools are

unable to reveal how zinc is stored and released, as these events are rare and hard to detect, says toxicologist Martin Philbert, who helped to develop the probes together with chemists Raoul Kopelman and Anne-Marie Sastry and their colleagues.

The potential of these and other nanoparticle tools is huge. James Baker, who studies such systems at the University of Michigan, points out that nanoparticles can be equipped with homing devices such as antibodies, reporting units such as tags that fluoresce when exposed to light, and even molecular systems for delivering drugs to a cell. Baker has, for example, used sphere-shaped polymers known as dendrimers to ferry a targeting molecule, fluorescent dye and methotrexate — a drug that attacks certain types of cancer cell — into a cell. In laboratory studies using tumour cells, the methotrexate killed 100 times more cancer cells when it was delivered by nanoparticles rather than simply being added to the cell culture⁴.

Gold prospect

Taking Heath's video-game analogy to its logical conclusion, nanotechnology could one day allow researchers to control the cell rather than just navigate around it. MIT bioengineers Kimberly Hamad-Schifferli and Joseph Jacobson can, for example, control the activity of proteins and DNA using gold nanoparticles. First, they embed the particles in either a protein or a DNA strand. Then they apply a radio-frequency magnetic field to a sample containing the nanoparticles and biomolecules, which causes the gold particles to heat up and disrupt the proteins' activity. Removing the field restores the proteins' function.

Last year, the pair used the technique to prise open a strand of DNA that had been twisted into a hairpin shape⁵. The system has not yet been used in cells, but the experiment was an impressive proof of principle. Other groups have used light-sensitive molecules to control protein function, but light cannot travel as far through living tissue as a magnetic field can. "We wanted something that



Anne-Marie Sastry and her colleagues aim to use tiny fluorescent probes to shed light on brain disease.

could be used in a really complex system like a cell or an organism," says Hamad-Schifferli.

So how good will these nanogadgets be at generating useful data for biologists? As more and more physicists become interested in these tools, there is a risk that they will churn out cool devices that seem great in principle, but have little practical use. "Until we get all the bugs worked out, a big challenge will be getting biologists to buy into it," acknowledges Carl Batt, a food-science researcher and co-director of the Nanobiotechnology Center at Cornell University in Ithaca, New York.

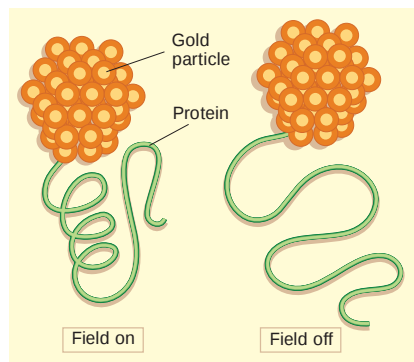
Basic questions, such as the degree to which intracellular probes disrupt the workings of the cell, remain unanswered. "How do you know you aren't measuring a physiological reaction to having a piece of junk inside the cell?" asks Batt. Adding to the scepticism is the phenomenon of nano-fatigue. "The 'nano' word is over-used and over-hyped," says John Ryan, director of the Nanobiotech-

nology Interdisciplinary Research Collaboration at the University of Oxford, UK.

The close partnership between physicists and biologists in the Alliance for Nano-Systems Biology does, however, suggest that some of these fears are being addressed. And most researchers contacted by *Nature* were excited by this coming together of disciplines. As nanotechnology brings more tools to the biologist's bench, Ryan asserts, the divisions between the fields of science will begin to break down. Universities are also helping by making courses more interdisciplinary. The approaches to problems may vary between biologists, chemists, engineers and physicists, but their common interests far outweigh their differences. "At the molecular scale," says Ryan, "all these fields reduce to the same basic science."

In this case, the science concerns the fast-paced activities of daily life in the cell. And for those in the nanosystems alliance, nanotechnology is the best way to get a grip on the many fleeting processes involved. Alliance member Leroy Hood, a molecular biologist at the Institute for Systems Biology in Seattle, predicts that nanotechnology will reveal as much new information about the cell as did the automated DNA sequencer — a device that he invented. "The combination of microfluidics and nanotechnology," Hood asserts, "will transform how biologists do everything." ■

Catherine Zandonella is a freelance writer in New York.



Kimberly Hamad-Schifferli (right) hopes to control proteins by attaching tiny gold particles to them — in a radio field the particle heats up, altering the protein's structure and inactivating it.



1. Melosh, N. A. *et al. Science* **300**, 112–115 (2003).
2. Vo-Dinh, T. J. *Cell. Biochem.* **87**, 154–161 (2002).
3. Klarreich, E. *Nature* **413**, 450–452 (2001).
4. Quintana, A. *et al. Pharm. Res.* **19**, 1310–1316 (2002).
5. Hamad-Schifferli, K., Schwartz, J. J., Santos, A. T., Zhang, S. & Jacobson J. M. *Nature* **415**, 152–155 (2002).

Alliance for NanoSystems Biology
♦ www.nanosysbio.org

Back from the dead

Under successive military governments, the discipline of forensic medicine nearly perished in Brazil. But it is starting to bounce back, inspired by a remarkable institute near São Paulo. David Adam pays the centre a visit.

It's harvest time in Brazil, and the sugar-cane fields are about to yield a bumper crop. But alongside the sugar will come a grisly haul. Fields of fast-growing sugar cane are a favourite dumping ground for killers with a corpse to hide. And when the harvest is gathered in, so too are the remains of missing people.

The task of investigating these discoveries falls to experts in forensic medicine. But in Brazil, these are in short supply — a legacy of the police state that prevailed from 1964 to 1985. Under military rule, forensic labs were directly controlled by police chiefs, which ensured that the activities of government torturers and death squads were covered up. Independent work in the field was all but impossible. And although the dark days of the 'disappeared' are now past, Brazilian forensic medicine remains largely moribund.

Except, that is, in Ribeirão Preto in the state of São Paulo, where a new centre for forensic medicine — termed legal medicine in Brazil — has been built on the fringes of the University of São Paulo's campus. Within its brightly painted walls, more than 300 kilometres north of São Paulo city, forensic pathologists work in relative independence, and in partnership with university researchers.

The centre owes its existence to the tenacity of its director, a quietly spoken forensic pathologist called Carmen Cinira Santos Martin. "It took 19 years of fighting," she says. Martin has been at the Ribeirão Preto medical school since 1980, and witnessed the decline of its legal-medicine department. Staff drifted away, and the building was eventually turned into an administration block.

Martin cut a lonely figure in those days: consigned to two shabby rooms in the corner of the pathology department with no laboratory facilities, her discipline became an option that tutors threatened their students with to ensure that they did their homework. "My teachers told me there was no future in it," says Sergio Britto Garcia, a pathologist who now works in the new centre for one day a week. "When you look around this modern building, you think that the university officials must have had a plan, but it wasn't like that. Carmen did it all. She's a hero."

Martin is modest about her achievement. "If I was alone, that meant I was the only one



who could do it," she says. Her opportunity came with a 1998 decision by the state of São Paulo to wrest its legal-medicine institutes from police control, making them accountable to state officials. "When I heard about the change in the law, I knew it was now or never," Martin says. Aided by another legislative change that encouraged partnerships between state institutions — in this case her university and the state officials newly placed in charge of legal medicine — Martin convinced the university authorities to provide US\$600,000 to construct the new centre.

It opened late in 1999, offering a new home for the 15 or so forensic pathologists of the former police institute, which was little more than a ramshackle hut in the city's cemetery. There,

On the case: Carmen Cinira Santos Martin (above) and Marco Guimarães are wrestling control of forensic medicine from the police chiefs in Brazil.



Grim harvest: fields of sugar cane are favoured by killers as a dumping ground for their victims.

break-ins were common, and gruesome pictures of murder victims made regular appearances in the region's tabloid newspapers. Today, post-mortems take place in the secure environs of Martin's new centre.

The arrangement benefits both its handful of academic scientists and the state forensic pathologists. "We now have a new model here in which we combine the research, modern technology and open mind of the university with the rich materials provided by the police," Martin says. "This is unique in Brazil."

For her part, Martin is investigating differences in the physical appearance of bones broken before and after death. That quest began when a skeleton was brought to the centre with a noose around its broken neck, but pathologists couldn't decide if the rope had been placed there before or after the injury was sustained. Was it suicide, or murder? Martin's work on the subject will appear later this year in the *Journal of Forensic Sciences*.

The eyes have it

Upstairs from the post-mortem rooms, pharmacologist Bruno Spinoza De Martinis is developing new analyses to detect the presence of illegal drugs and alcohol in a dead body. With Martin, he will shortly publish his first paper: a test for alcohol using samples of vitreous humour from the eyes. The technique is needed in Brazil because the hot, humid climate quickly rots corpses and alters their blood chemistry, De Martinis explains, proudly pulling a pre-print from a drawer stuffed with evidence bags containing LSD and ecstasy. Originally a chemist studying atmospheric pollution, De Martinis decided to retrain to work at the centre after meeting Martin at a Christmas party. "I wasn't sure about working with dead people, but Carmen talked me into it," he says. "She's very persuasive."

Another researcher who succumbed to Martin's powers of persuasion is Marco Guimarães, a pathologist who is now leading efforts to develop a laboratory for human

identification. Guimarães spent most of last year in Britain learning forensic DNA techniques at the University of Sheffield, and is now applying this expertise to help solve one of Brazil's most enduring forensic mysteries. In collaboration with scientists in São Paulo city, Guimarães is working to identify skeletons recovered from a clandestine cemetery, some of which are believed to be those of dissidents murdered by government death squads.

The cemetery was discovered in 1990 on the outskirts of São Paulo, and contained the remains of up to 1,200 people. Most would have been poor residents or vagrants, but at least 15 of the skeletons are thought to belong to the ranks of the disappeared. Martin and Guimarães hope that the Ribeirão Preto centre will help to lay their ghosts to rest.

The skeletons were originally sent for identification to the legal-medicine department at the nearby University of Campinas. There, a team under the supervision of Fortunato Badan Palhares, one of Brazil's foremost forensic scientists, used facial-reconstruction techniques to match the remains with photos of the dissidents while they were alive. Progress was slow, and after articles in the press alleged that Palhares was linked to police and state corruption, the skeletons were returned to São Paulo in 1997.

They are now in the care of forensic pathologist Daniel Muñoz, who works jointly in the city's legal-medicine institute and the University of São Paulo. His pride stung when the skeletons were sent to the rival Campinas centre, Muñoz is determined not to let them go. "We can do it here in São Paulo," he insists, pointing to his team's success in identifying the passengers of a plane that crashed in São Paulo in 1996. But Muñoz has enlisted the help of Guimarães, who has already made a preliminary match between DNA from a tooth from one of the skeletons and a sample taken from a close relative of a political activist last seen alive in 1972.

Such work would be almost impossible in police institutes, which are forced by their meagre facilities to contract out most DNA identification work. This is an expensive

A crowd holds up pictures of television journalist Tim Lopes, who was killed last year.



business, so cases are often only investigated if a third party foots the bill. Last year, for example, Franklin David Rumjanek, a biochemist at the Federal University of Rio de Janeiro, used DNA techniques to identify charred bone fragments suspected to belong to television journalist Tim Lopes, who disappeared in June 2002 while investigating drug trafficking and child prostitution in one of the city's slums. Lopes' employer, the media company TV Globo, paid for the analysis. If it hadn't, the five suspects now awaiting trial for his murder would probably still be at large.

Remains of the day

In São Paulo, Muñoz finds his ambitions similarly frustrated. Several hundred skeletons a year arrive at his state legal-medicine institute for identification, but most end up being buried without ever being analysed. One recently arrived example is lying on the table in his lab. Found at the side of the road, the body was burned, and its charred clothes and a melted pocket radio are piled next to it. One thighbone has been cut into, to help judge the age of the victim. "We can work out the sex, height and age in most cases, but not much more," Muñoz says.

The skeleton will soon be moved to a room down the corridor. "Prepare yourself, we don't yet have all the furniture we need," says Muñoz, unlocking its door. From floor to ceiling around the walls, human bones are piled in plastic boxes, and bags containing more skeletons cover the floor.

Back in Ribeirão Preto, Martin's new centre, set in the rolling grounds of a former coffee farm, seems a world away from the urban sprawl of São Paulo and Muñoz's chamber of horrors. Martin says that this distance from the state's capital, where some politicians are still suspicious of the idea of setting forensic experts free from police control, is one reason her centre has managed to win a degree of independence for its scientists. Still, delaying tactics in the state parliament mean that the official paperwork endorsing the creation of Martin's new centre has not yet been signed. And money for equipment and reagents remains extremely tight.

Martin won't let these obstacles hold her back, but whether one small institute can inspire a renaissance of forensic medicine across such a huge country remains unclear. Nonetheless, Tim Cahill, Brazil specialist for the human rights charity Amnesty International, believes that Martin has established a welcome precedent. "It's a very important step to directly link the forensic medics with the university," he says.

At the very least, the centre's work should mean that the fields around Ribeirão Preto will become a much less welcoming place for killers who want to ensure that their victims are never identified.

David Adam, until recently a news and features writer for *Nature*, now writes for *The Guardian* in London.

Improving science through online commentary

The Internet offers a timely opportunity to widen, and reduce delays in, scientific debate.

Sir—Progress in science depends on public debate and criticism of ideas. Unfortunately, debate is largely restricted to four domains: conferences, private conversations, journal clubs and peer-reviewed publications. The first three channels tend to be too private and ephemeral to help the community at large. Publication in journals runs on a timescale of months and, unless people bring new data, they are lucky to be allowed into the debate at all. This slow communication hampers progress. Most people do not know other scientists' views on published manuscripts. When a new paper appears, readers often spot logical flaws, experimental weaknesses, questionable assumptions or alternative interpretations. Yet individual criticisms may not be considered important enough to warrant publication. Even major criticisms are unlikely to appear until months or years later, and are often overlooked in the haystacks of the literature. Scientists even miss official retractions, continuing to cite withdrawn data accidentally.

We propose a simple solution. Each record in publication databases (such as the US National Library of Medicine's PubMed) should have a link for adding scientific commentary: essentially the electronic version of a Post-it note. The submitter of a comment would be required to provide his or her name, e-mail address, a short communication and potentially a password (for later editing or deletion). Insights, criticism, replications and non-replications could be posted.

The benefits of this system would include: immediate free and open debate of scientific ideas and results; easy dissemination of non-replications and negative results; a reduction in wild-goose chases by researchers unaware of others' insights into a published paper; and posting of links to subsequent papers that provide important verifications or contradictions.

The attachment of commentary to databases has been successful in other areas. Users can add 'Post-em' links to a public-domain genealogy database, the Social Security Death Index (<http://ssdi.genealogy.rootsweb.com/cgi-bin/ssdi.cgi>). Readers attach reviews to books in Amazon.com's database.

Since we desire our proposal to be practical and inexpensive, there would be no human gatekeeper. To avoid abuse, we suggest that unhelpful or libellous notes could be reported to a removal moderator, who would decide whether to eliminate them. To prevent milder forms of abuse, Amazon.com asks: "Was this review

helpful to you?" Answers are tallied to give a sense of how the community values the comment. Automatic advertisement posting can be prevented by a requirement to key in a text code readable only by humans.

Given the importance of peer-reviewed publications to their careers (see P. Lawrence *Nature* **422**, 259–261; 2003), will researchers freely provide their insights without any career benefit? They already do. During the peer-review process, most scientists spend a lot of time providing insights to a very few people. Posting a note would benefit the whole community and take very little time.

A handful of journals have begun experiments in 'open' peer review online (see T. Gura *Nature* **416**, 258; 2002). This operates on a journal-by-journal basis, whereas our proposed post-publication commentary process should be centralized

and independent of the journal of publication. For example, the commentaries on biological and medical papers should be stored in a single database under the free-access, disinterested aegis of the US National Library of Medicine.

We believe that this simple, practical idea could substantially speed the practice and progress of science. We address further issues and welcome feedback at www.cnl.salk.edu/~eagleman/postit.

David M. Eagleman*, **Alex O. Holcombe†**

*Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA
eagleman@salk.edu

†Department of Psychology, University of California, San Diego, La Jolla, California 92093, USA
holcombe@post.harvard.edu

Oil and war: we had the warning 30 years ago

Sir—The other night, I was looking for an article in *Nature* from 1973 when the volume fell open to an editorial, "What future for energy?" (*Nature* **242**, 357–358; 1973). I quote the closing part: "...there is every prospect that the immediate problems of energy scarcity could be brought within control in the measurable future. Fuller use of the price mechanism and more resolute decision-making by the Federal Government are urgently necessary, but could make the immediate "crisis" go away. By the 1990s, however, the United States will have more serious problems to contend with. It now seems clear that the annual import bill for petroleum products is bound to increase steadily, and the chances are that even by 1980, the United States will be spending \$20,000 million a year (or more if there are further devaluations of the dollar) on imported crude petroleum. Numerically, this expenditure is not insupportable, but there are political snags. The Administration is, for example, alarmed at the prospect that so much cash will find its way into the hands of Arab states which have no automatic sympathy with United States policy in the Middle East — rather the reverse — and which could use their holdings of the large quantities of dollars as a way of manipulating the international currency market. But even this is not an entirely unwelcome prospect, for at the very least it will create a climate in which some lasting political settlement in the Middle East will seem to

everybody desirable. And here, luckily, there is plenty of time in which the United States can prepare the ground for a lasting policy. To be sure, in the meantime it will be prudent also to pursue as vigorously as possible the technical means by which other sources of energy might be exploited, and no doubt there is much to be done to make use of geothermal steam and solar energy. But the problem of the 1990s is already definable and is largely political in character. It is to be hoped that the United States will not let that blow up into a crisis for lack of resolution in the next few years."

Thirty years later, the Israeli–Palestinian situation could not be worse. War in Iraq has cost thousands of lives during the past few weeks. Republicans in the US Congress recently fell just two votes short of passing legislation that would allow oilfields to be drilled in the Arctic National Wildlife Refuge on Alaska's North Slope, and the president vowed to redouble efforts to open the refuge to drilling. There have been no significant investments in new energy technologies, and no breakthroughs in technologies or new sources.

The perspective offered by your stunning, telling editorial in 1973 provides some important, albeit expensive, lessons.

Michael D. Jennings

Donald Bren School of Environmental Science and Management, University of California, Santa Barbara, and Department of Fish and Wildlife Resources, University of Idaho, PO Box 441136, Moscow, Idaho 83844-1136, USA

The author of the editorial was John Maddox, then Editor of *Nature* — Editor, Correspondence

The buck stops here

Do we have free will, or are all of our choices predetermined?

Freedom Evolves

by Daniel Dennett

Allen Lane: 2003. 352 pp. £20

Viking: 2003. \$24.95

Melvin Konner

The puzzle of free will is not something new. Intelligent people have always wondered how, if all is fated, or if an omniscient God can know all in advance, choice can be free. Even a child can see the problem: if God knows what I am going to do, how can my choice be meaningful? Theologians have answers to the child's question, of course, but none of them satisfies scientists.

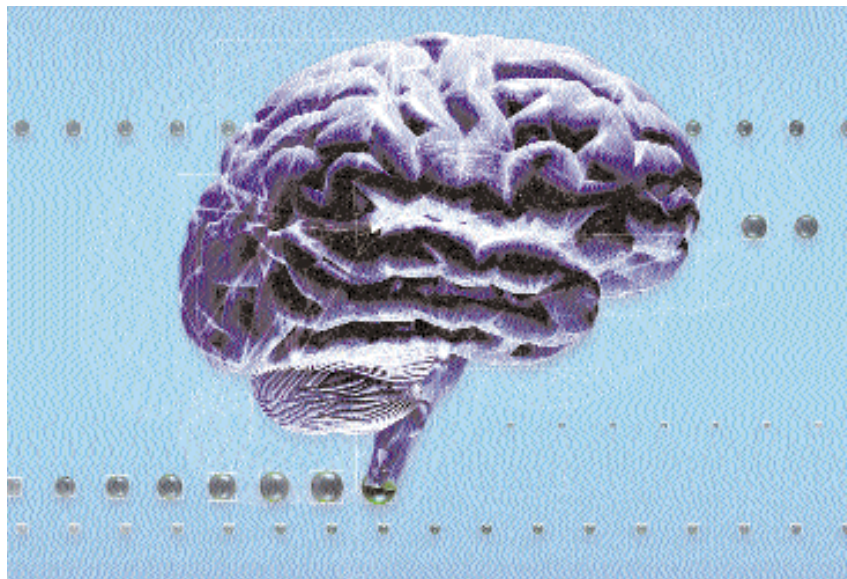
The best answer, perhaps, is that God does not direct your choice but merely knows what you will do with it. This still leaves the paradox of predestination, but religious people who believe in an omniscient God cannot escape the subjective sense that they are making choices.

Scientists have a parallel, possibly more difficult, problem. Our minds, we believe, are instantiated in our brains, and our choices are at the end of a causal chain that includes environmental influences operating on a developmental process resulting from a coherent, hierarchically ordered set of at least 30,000 genes. This set of genes is in turn the result of natural selection operating imperfectly on previous sets of genes over several billion years. Behind those first modestly stable strings of nucleic acids were another ten billion years or so of physical and chemical events going back to the Big Bang, which I call Lumina.

Lumina's cause is indeterminate, but every single event that followed it, including my decision to write this review or yours to read it, has been in an important sense determined. Quantum indeterminacy is about what cannot in principle be known, but each step leading up to a choice involves so many quantum events that we can sum them statistically and ignore quantum indeterminacy. So my choice to write the review was determined by the time I answered the invitation. In what sense, if any, then, can my choice have been free?

Daniel Dennett undertakes to answer this question decisively in this engaging and very smart book. As in his previous books, about the nature of consciousness and the implications of Darwin's work, for example, he has rendered the difficult comprehensible and has ably defended a clear piece of intellectual territory. If a main aim of philosophy is the clarification of discourse, then Dennett is extremely good at his trade.

He is part of a growing group of philoso-



Free thinking? Decisions made in the brain are at the end of a causal chain going back to the Big Bang.

phers, beginning with David Hume in the eighteenth century (although Aristotle had related leanings), who believe that they must understand science. They believe this not because they think that empirical facts drive philosophy, nor because they want scientists to lead them around by the nose, but because they think the most profound argument cannot stand unless it is consistent with the facts, and science is the best way to discover facts.

Dennett eliminates certain old arguments about the mind by demonstrating their incompatibility with facts: for instance, that neural circuits are the entire basis of consciousness; that there is no cartesian theatre in the brain; and that evolution explains all of life, including ourselves and our moral choices. But here his goal is more specific and grander. He proposes to show not only that such facts are no threat to free will, but that they positively enhance it.

I can't do justice to his claims here, but I did not find them convincing. In brief, he takes the position that free will and determinism are compatible, and that those who think they must relinquish one or the other are yielding to a merely imagined constraint. But most of the book consists of conventional arguments in favour of determinism itself, which alone cannot support his view.

It is true, for example, that the mind is not separate from the brain, but this fact cannot enhance free will. It is true that evolution has enlarged our freedom of choice; "free as a bird" is lovely but is no compliment to a person who has at her disposal the power of culture and history. Yet such a person,

like the bird, is constrained by genes, habits, energy, gravity and other forces. It is true that psychologists can easily insert false beliefs into decisions, and that conscious choices are temporally smeared or spread out in the brain, but how do these facts free us? No one denies that we are unique in the animal world in raising these questions, but do we do so by choice? The puzzle of free will, I think, remains unsolved.

Nevertheless, the book is wonderfully interesting. Scientists may know many of the themes covered here — the Life game's simulation of evolution, the nature of neural-network models, the blindness of natural selection — but they may be news to some philosophers, who will need to grapple with their implications. As for the general reader, and I hope there will be many, the book is a succession of powerful ideas explained in lively and highly original ways.

Dennett's demolition of bad arguments is entertaining. The philosopher John Austin argued that, while he kicked himself for not having made a golf putt, under those exact conditions he could not have done better. Dennett counters that what we mean by "I might have done it" refers not to those exact conditions but to a set of roughly similar conditions. Robert Kane advanced a theory of "self-forming actions", occurring at critical junctures in life when our choice is not determined — breaks in the causal chain. But why would those choices escape causation more than any other?

Yet there are, I think, two partial refuges for free will at the end of the causal chain of

evolutionary, genetic, developmental, environmental and neural determinants. The first is that this chain is almost completely unknowable. There are so many steps, each of which is so sensitive to initial conditions (defined at any arbitrary starting point) that a computer the size of the Universe, with as many transistors as there are quarks, and all the time remaining until entropy reigns could not predict more than probabilistically the simplest human decision.

The second refuge is more important. My mind may be a survival machine with pre-determined choices, but I live inside it. I can embrace the facts of neurobiology, while rejecting the notion of using it to convey my subjective experience. "I was walking on a dark street and the central nucleus of my amygdala reported signals to the basolateral nucleus." No. "I was afraid." You immediately know what I mean because you have been afraid, and the unique human capacity for language-based intersubjectivity allows you to understand me.

I experience choice the way I experience fear, and you experience my choices — praising some, punishing others — as you experience my fear. It is simply a misuse of language to say that the causal chain threatens freedom, because what we mean by free will is part of our subjective experience. We may be Lumina's children, but she does not control our choices, because what we mean by 'choice' is not freedom from the causal chain; it is our insider's view of one link, as vital and inexplicable as fear, joy or love. I trust that you will hold me responsible for my choices, regardless (with few exceptions) of their many unknown causes. We live, you and I, in a subjective world of choices, and that is all we really need to know about free will. ■

Melvin Konner is in the Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA. He is the author of the revised edition of *The Tangled Wing: Biological Constraints on the Human Spirit*.

Secrets of the tomb

The Scientific Study of Mummies
by Arthur C. Aufderheide
Cambridge University Press: 2003. 626 pp.
£100, \$150

Sarah U. Wisseman

The esoteric discipline of mummy studies has come a long way in the past 30 years. In the 1970s, scientific teams at Manchester Museum in Britain and the University of Pennsylvania in the United States undertook major projects that included full autopsies of Egyptian mummies, gaining valuable information about lung and heart diseases, dental



The mummy returns: the remains of Tollund Man were unearthed after 2,000 years in a Danish bog.

health and intestinal parasites, as well as the fabrics and fluids used by the embalmers. Needless to say, the subjects of these autopsies lost their value as museum display objects. Non-invasive techniques, such as X-ray radiography and computed tomography (CT), have subsequently been used to achieve some of the same objectives without harming the exterior of the mummies.

The Scientific Study of Mummies is a veritable encyclopaedia of mummies and of mummy-research projects. Arthur C. Aufderheide is uniquely equipped to write such a book, having personally dissected more than 500 mummies. First he sets the stage with the history, purpose and mechanisms of mummification. Then he takes us on a world tour of mummies, ranging from the Americas (including the Peruvian Ice Maiden and a spontaneously mummified woman dubbed Miss Chile) to Europe (the Alpine iceman and Danish bog bodies, along with many less-well-known mummies), the Canary Islands (home of the Guanche mummies and the site of the First World Congress on Mummy Studies in 1992), Egypt, China, Japan, Australia and New Zealand.

The second half of the book reads almost like a medical text and assumes a basic knowledge of chemistry. The section on the soft-tissue taphonomy of mummies, defined by the author as "the effects of postmortem processes that altered the decay mechanism sufficiently to result in mummification", is quite technical, but is clearly presented with ample illustrations. This discussion features numerous agents that cause decay and deterioration: water, bacteria, fungi and temperature fluctuations. Aufderheide then describes the full range of invasive and non-invasive research methods that have been used by mummy researchers, from dissection and tissue sampling to radiology, electron microscopy, endoscopy and magnetic resonance imaging (to be successful, the last technique usually requires the rehydration of

the mummy tissues). In the palaeopathology section, we learn about the diseases and abnormalities that afflicted ancient as well as modern humans: everything from ear infections to silicosis. Lung tissue is often better preserved than other parts of the body, so there is more information on lung diseases than on other types. The author also describes the parasites found in mummies, such as the fish tapeworm that infected people who ate uncooked fish. As he notes: "the biomedical data these mummies supply tells us how the diseases we presently suffer have evolved".

The author's training as a pathologist is reflected in his content and emphases. His discussion of Egyptian mummification chronology stresses the physical aspects of the mummies rather than their historical contexts. Cross-comparison of materials used in embalming, arm and leg positions, packing and organ removal are useful in understanding the range of practices over time, but no chronology can be complete without the archaeological context (such as site and tomb if known, region and circumstances of recovery). Another problem, as Aufderheide himself points out, is that the data he uses range from observations made 100 years ago to recent work. One must also take into account the changes in imaging resolution over time; our view of many mummies would benefit from re-examination by modern X-rays and CT scans.

The extensive bibliography includes both medical and archaeological references. A glossary of technical terms would have been helpful, particularly for those without training in the biological sciences. Nonetheless, this book should be required reading for any serious researcher on the sometimes gory but always fascinating subject of mummies and what they can teach us about human beliefs and behaviour. ■

Sarah Wisseman is in the Program on Ancient Technologies and Archaeological Materials, University of Illinois, Urbana-Champaign, Illinois 61801, USA.

The pathology of history

The Deadly Truth: A History of Disease in America

Gerald Grob

Harvard University Press: 2002. 368 pp.

\$35, £23.50, E35

W. F. Bynum

A generation ago, historians discovered the importance of disease in human history. *Plagues and Peoples* (1976) was written by William H. McNeill, then professor of history at the University of Chicago, who had previously provided admirable synoptic accounts of world history. His book was a revelation, even to medical historians, who were then concerned with doctors rather than patients or disease. Gerald Grob, a distinguished historian of psychiatry and medicine, cites McNeill's book in his own more modest account of disease in American history.

Grob has always been known for his quiet medical realism, even in the problematic field of psychiatric history. This makes his sombre reading of the past, present and probable future history of disease in American society all the more telling. There is no social constructivism here, rather the more convincing historical insight that geography, genetics, behaviour, social perceptions and economic realities have shaped disease realities in the past, and are likely to continue to do so. Indeed, the real guiding light behind Grob's analysis is not McNeill, but the microbiologist René Dubos of the Rockefeller Institute (as it was then called), who told us more than four decades ago that medical utopias are mirages. Dubos would have applauded medical advances since, but his conclusions, I believe, would have been much the same.

As a historian, however, Grob's brief is with the past, not the future. He takes a long view in a short book, from the consequences of European settlement for the health of native Americans, to the age of AIDS and other chronic diseases. He offers several particular insights and three broad themes that are worth pondering.

First, geography (and hence climate) has shaped the history of disease in America. Massachusetts, Virginia and South Carolina posed radically different challenges to the various groups of Europeans who tried to settle there. Hard winters often meant food shortages with attendant nutritional disorders and diseases of overcrowding. The American south had its own health problems, including the insect-borne scourges of malaria and yellow fever. Smallpox was no respecter of climate, although it was a convenient agent in the European struggle with the native inhabitants. Infected blankets were sometimes traded to unsuspecting

Art

A leap into the future

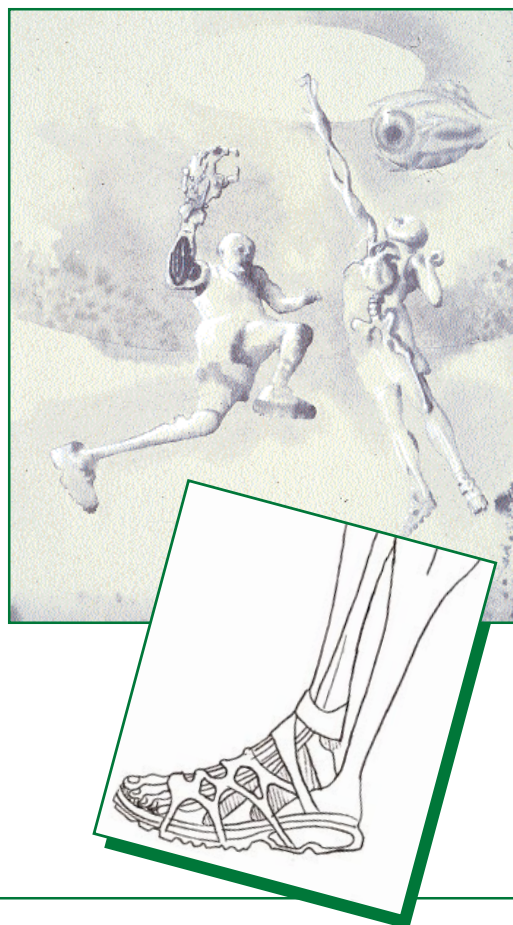
How far can science and technology push the frontiers of human sporting achievement? This question occupied New York artist Alexis Rockman during his year-long residency at Camden Arts Centre in London, which has just ended.

The studies shown here for his work in progress, *Future Olympics*, show the fruits of his long discussions with physiologists and molecular biologists in London and Oxford. They display his signature motif: the grafting together of the artificial and the organic, which has been a hallmark of much of his earlier work.

This style was evident in Rockman's famous image *The Farm*, which led to his portrayal as a harbinger of the dangers of genetic engineering. He confesses that his view of science is "confused", believing that science may have caused problems but that it will also help to solve them.

Rockman will take part in 'Superhumans and Superstitions: The Contested Identity of the Human Body', a public panel discussion to be held at the Royal Institution in London on 8 May.

♦ www.rigb.org



CORNEY, BRAVIN & LEE

tribes; bioterrorism is no new phenomenon.

The south continued to bear a geographical health burden until well into the twentieth century, but in nineteenth-century America, the cities, wherever they were located, were central to the geography of health. New York, Philadelphia and the other cities on the east coast received the bulk of immigration from Ireland and southern and eastern Europe. Diseases of deprivation and destitution were rife, and sanitary reform was slow and partial. What demographers call the urban penalty relaxed only gradually, and still exists in a variety of forms in cities everywhere.

The second theme of Grob's monograph is the continuing importance of gastrointestinal disorders. Cholera has received a lot of historical attention, but Grob reminds us that dysentery and other diarrhoeal disorders were a far more common cause of morbidity and mortality than the more spectacular cholera epidemics. Infants and children were particularly susceptible to gastrointestinal afflictions, even before the increase in bottle-feeding during the nineteenth century. Like the diseases themselves, the literature on cholera and consumption is more visible historically. But diarrhoea

and its consequences caused concern, and historians ought to continue to explore this cluster of diseases.

Grob's final theme is that disease is not about to disappear. His book is properly appreciative of the contributions of modern biomedicine to the health and longevity of Americans and other fortunate citizens of the Western world. But he writes from the sombre perspective of the new millennium, when the problems of modern medicine tend to be more newsworthy than its achievements. Talk about the conquest of acute infectious diseases now seems hollow, and we dwell instead on the chronic ones that have replaced them. A generation ago, Dubos was a prophet crying in the wilderness. Grob is a chronicler of his times.

Grob's achievement is considerable. He has read and synthesized a massive literature on the social history, demography, and medical and cultural experience of a large country. His judgements are judicious and his book deserves the ultimate compliment: it would have been better had it been twice as long. ■

W. F. Bynum is at the Wellcome Trust Centre for the History of Medicine at University College London, 24 Eversholt Street, London NW1 1AD, UK.

The unsung heroes

Peter Parham

Immunology ostensibly began with ancient Greek physicians, who marvelled at their observation that patients who had survived a bout of plague were resistant to the disease when it came around again. From this germ was born the idea of vaccination — a small controlled infection, like a sort of immunological rehearsal, designed to confer future resistance to full-blown infection while causing only mere discomfort, rather than disease. What those Greeks also initiated was the tendency for immunologists to tackle their subject in a backward manner. They had become captivated by immunological memory, a late-acting feature of the immune response that only becomes useful if and when the first battle between person and pathogen — the primary infection — has been won by other means.

Memory is the end point for adaptive immunity, the part of defence that is mediated by white blood cells called B cells and T cells. Although fascinating to immunologists for their strength and specificity, these soldiers are not speedy. Typically found in a state of suspended animation, B and T cells only become fit to fight about a week after combat begins. During this first week of a primary infection, defence is in the hands of innate immunity. One of several paths can be followed: the infection can gain the upper hand, leading to death or chronic, debilitating disease; alternatively, an acute infection can cause a temporary state of disease, which ends when innate immunity calls

up reinforcements in the guise of adaptive immunity. Patients who experience either of these outcomes have been carefully studied by physicians ever since the time of the ancient Greeks.

A third possibility is that the infection will be terminated quickly by innate immune mechanisms without the involvement of adaptive immunity and with no major symptom of disease. In this happy circumstance, in which health is maintained or only mildly perturbed, neither ancient Greeks nor their modern counterparts would be likely to consult a physician. Consequently, this most desirable of outcomes, in which the war is won with little collateral damage, remains understudied and poorly appreciated.

The fact that innate immune mechanisms can terminate infections should not be surprising — invertebrates and jawless fish survive infections solely on the wits of their innate immune systems. Even mutant laboratory mice with no adaptive immunity can survive and reproduce, as long as they are kept apart from their non-mutant brethren. Given this potential, what proportion of infections is terminated by innate immunity? In modern human society, we seem to endure intense and relentless exposure to all manner of infectious agents. Perhaps the very fact that most people are not perpetually sick is testament to innate immunity squelching most of the infections that we contract.

If we assume that innate immunity does, with some frequency, terminate infections before the onset of disease, there should also be natural selection for variants of proteins in the innate immune system that increase that frequency, even if only temporarily. Conversely, polymorphism among proteins of innate immunity is evidence of selection for such improvements in innate immunity. Several examples of this phenomenon have been revealed, such as the receptors and ligands that control the functions of natural killer (NK) cells, the lymphocytes of innate immunity.

NK cells are large, well-armed and circulate in a state of readiness, enabling them to enter and defend a tissue almost as soon as it becomes infected. NK cells kill infected cells and secrete cytokines, which rally the other cells of innate immunity, such as microbe-masticating macrophages. Moreover, through interaction with dendritic cells, NK cells help to decide if and when an adaptive immune response is needed. When innate immunity extinguishes an infection unaided, NK cells may kill off antigen-bearing dendritic cells, preventing their migration to lymphoid tissues where they would otherwise activate T cells.

Innate immunity

The fact that most people are not perpetually sick is testament to innate immunity squelching most of the infections that we contract.

Unlike B and T cells, NK cells forego the process of rearranging genes to produce a repertoire of antigen receptors. Instead, they assemble different combinations of receptors from an impressive selection of activating and inhibitory cell-surface receptors. Many of these receptors bind specifically to major histocompatibility complex (MHC) class I or class I-like molecules — ligands that also control the activity of killer T cells and were previously thought to be dedicated to adaptive immunity. Prominent among receptors on NK cells are the killer-cell immunoglobulin-like receptors (KIRs), which, like their MHC class I ligands, are encoded by a diverse and rapidly evolving family of genes. People differ in the number and type of KIR genes that they inherit; the NK-cell repertoires of all but the closest of relatives are distinct. We are beginning to see how this individuality in NK-cell receptors is correlated with varying strength and weakness in the response of these cells to infection.

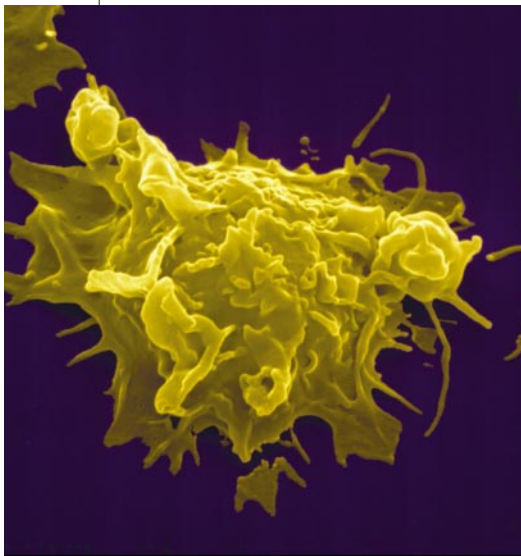
Innate immunity has sometimes been viewed as ancient in origin, and therefore highly optimized by natural selection. But does this mean that it is now set in evolutionary stone? Although the mechanisms of innate immunity predate those of adaptive immunity, it is not necessarily true that these mechanisms have stopped evolving. NK cells in humans and other species have many types of receptor, some conserved and others highly variable. These are not the properties of a highly optimized system, but rather of one that has been successively pushed in many different directions by natural selection. The fact that innate immunity is not unchanging, but rather is plastic and continuing to evolve, can be readily understood if we realize that innate immunity has the capacity to prevent primary infections from actually causing disease, an attractive feature that is out of bounds for adaptive immunity. ■

Peter Parham is at Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305, USA.

FURTHER READING

Parham, P. (ed.) *Immunol. Rev.* **182**, 1–289 (1999).
Martin, M. P. *et al. Nature Genet.* **31**, 429–434 (2002).
Scalzo, A. A. *Trends Microbiol.* **10**, 470–474 (2002).
Moretta, A. *Nature Rev. Immunol.* **2**, 957–964 (2002).

EYE OF SCIENCE/SPL



Attacking defence: a natural killer cell is among the first troops to enter immunological battle.

Life's transistors

Fred J. Sigworth

Voltage-gated ion channels control electrical activity in nerve, muscle and many other cell types. The crystal structure of a bacterial voltage-gated channel reveals the astonishingly simple design of its voltage sensor.

The membranes of living cells, from bacteria to humans, contain protein macromolecules that behave rather like field-effect transistors. In transistors, the flow of electrons through a semiconductor 'channel' is governed by the voltage applied to a 'gate' electrode. With the protein equivalents — voltage-gated ion channels — an appropriate voltage, imposed across the cell membrane, causes the channels to open and allows a current of ions to cross the membrane. The molecular structures within ion channels that sense the membrane voltage have remained obscure for the 50 years since Hodgkin and Huxley first described¹ their function. But the voltage sensors have at last been made visible, in the X-ray structure of a potassium ion channel. Youxing Jiang, Roderick MacKinnon and colleagues present this work on page 33 of this issue², and in a second paper (on page 42)³ they describe tests of a hypothesis for voltage-sensor motion.

The functional unit of a voltage-gated channel is an assembly of four proteins, or subunits; in each, the polypeptide chain snakes back and forth across the membrane six times. This 'six-transmembrane' structure is seen in the voltage-gated potassium, sodium and calcium channel families, and also in other channel types. As voltage-sensing devices, these channels can perform much better than their electronic counterparts (Fig. 1a). Their high sensitivity to voltage is important, because cellular voltage changes are small.

A simple biological voltage sensor would be a charged particle within a cell membrane, with an imposed voltage difference driving the particle from one surface of the membrane to the other. Theory shows that, to explain the observed sensitivity of voltage-gated channels, at least 12 elementary charges must participate in the sensing mechanism. Indeed, direct measurements of 'gating currents' confirm that the total charge displacement in a channel's voltage sensors is about 13 elementary charges per channel⁴.

So where in the channel protein are these charges found? It has been established⁵ that, in a six-transmembrane channel, the transmembrane segments known as S6 helices — one from each of the four subunits — form a 'gate'. That is, they form a bundle that can pinch off the ion pathway, effectively closing the channel. Meanwhile the fourth segment,

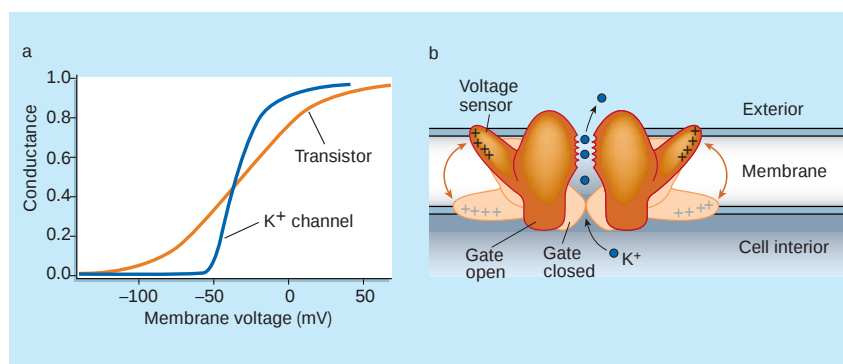


Figure 1 Voltage sensing in a potassium ion channel. a, The control of ion flow through voltage-gated channels is very sensitive to the voltage across the cell membrane. By comparison, an electronic device such as a transistor is much less sensitive to applied voltage. b, MacKinnon and colleagues^{2,3} have found that the voltage sensors in a bacterial potassium channel are charged 'paddles' that move through the fluid membrane interior. Four voltage sensors (two of which are shown here) are linked mechanically to the channel's 'gate'. Each voltage sensor has four tethered positive charges (arginine amino acids); the high sensitivity of channel gating results from the transport of so many charges, 16 in all, most of the way across the membrane.

S4, has always seemed a natural candidate for the voltage sensor that causes the gate to open and close⁶. Depending on the channel type, S4 has four to seven positive charges, mostly from arginine amino-acid residues. Moreover, S4 is otherwise very hydrophobic, which makes it likely to be embedded in the oily interior of a protein — or within the membrane itself. Because of these properties, S4 has been the subject of intense study⁷. Mutation analyses have shown it to be important in voltage sensing; spectroscopic probes have shown that it moves in response to voltage; and chemical-modification studies have revealed that some of its amino acids are alternately exposed on the internal or external face of the membrane, depending on the membrane voltage. So it has become clear that the four S4 segments per channel are the main voltage sensors.

What structural design would allow so many charges to move so far, crossing the 30-Å-thick, electrostatically hostile interior of a cell membrane? Practically everyone in the ion-channel field (including myself) has imagined the S4 segment to be an α -helix — a common structural feature of proteins, in which the polypeptide backbone is twisted into a spiral — that is packed snugly among the other helices of the protein. It has been thought that the S4 helix would undergo a shift or a rotation in response to voltage, and that charge transport might even be ampli-

fied by 'focusing' the membrane electric field near S4. This model has made its way into the textbooks. But the results of MacKinnon and colleagues show that it is almost certainly wrong.

Determination of the structure of a voltage-gated channel has been long in coming. The five-year effort in the MacKinnon laboratory involved trials of many channel proteins, none of which formed either two-dimensional or three-dimensional crystals. Reasoning that these failures might reflect a particularly loose protein structure, MacKinnon and colleagues decided to use parts of antibody molecules (so-called Fab fragments) as a scaffold to aid crystallization, and also chose a particularly rugged channel protein. Although its origin is the archaeobacterium *Aeropyrum pernix*, this protein, KvAP, has sequence features and electrical characteristics⁸ that place it firmly in the broad family of voltage-gated potassium channels.

The resulting X-ray structure² of KvAP shows the expected potassium channel core, consisting of transmembrane segments S5 to S6, surrounded by S1 through to part of S3. What was unexpected is that the S4 helix, along with the second part of S3, forms an α -helical hairpin — a 'paddle' that extends out from the channel core into the membrane's fluid interior (see Fig. 3 of ref. 2, page 35). The paddle has a flexible connection to the

rest of the channel, as the authors show by comparison with another crystal structure, of segments S1 to S4 alone. This flexibility explains the difficulty that the authors encountered in crystallizing the protein; it also suggests a mechanism for voltage sensing. The paddle is a hydrophobic, charged particle that can move in the membrane interior, transporting its four positive charges from one membrane surface to the other (Fig. 1b).

It is the location of S4 — not embedded in the protein core, but loose in the membrane — that is the big surprise here. It explains an old puzzle, that small lipid-soluble molecules somehow have ready access to ion-channel voltage sensors. Such molecules include local anaesthetics, the alkaloid nerve toxins and the well-known insecticides allethrin and DDT. It is now easy to imagine them diffusing up to the voltage-sensor paddle from within the lipid membrane interior.

An X-ray crystal structure is like a posed photograph; in the KvAP crystal, for instance, the voltage-sensor paddle is held firmly in place by an antibody scaffold. What can be learned about the paddle's natural conformation and movements? A few years ago, Horn and colleagues⁹ showed for sodium channels that a bulky moiety, attached by chemical modification to an S4 amino acid on the outside of the membrane, can actually be dragged through to the inner surface in response to an inside-negative voltage. In their second paper³, MacKinnon and colleagues show that a much larger molecule — biotin plus a 17-Å linker — flips across the membrane in a voltage-dependent manner when it is attached to an S4 amino acid in KvAP. They conclude that the S3–S4 paddle moves through a quite unrestricted space. They go on to attach this biotin–linker molecule to various other sites in the paddle, to map its position relative to the membrane surfaces at positive and negative voltages.

After all this, MacKinnon and co-workers have still left a few questions to be answered. The actual conformation of the channel in the membrane will need to be clarified, because in the crystal the membrane is replaced by a blanket of detergent molecules. Questions also remain about the disposition of the amino-terminal end of the protein (thought to be intracellular) and of the loop between the S3 and S4 segments in related channels (in the well-studied Shaker potassium channel, this loop is always accessible from the outside surface). Moreover, details of the motions of the voltage sensor — in some channels the charge movement occurs in several discrete steps — remain to be worked out, as does the energetic issue of moving the quadruply charged paddle through the membrane interior. But the structure of KvAP's voltage sensor, so simple and, with hindsight, so obvious, is a wonderful end to a 50-year-old mystery. ■

Fred J. Sigworth is in the Department of Cellular and Molecular Physiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520-8026, USA.

e-mail: fred.sigworth@yale.edu

1. Hodgkin, A. L. & Huxley, A. F. *J. Physiol. (Lond.)* **117**, 500–544 (1952).

2. Jiang, Y. et al. *Nature* **423**, 33–41 (2003).
3. Jiang, Y. et al. *Nature* **423**, 42–48 (2003).
4. Bezanilla, F. *Physiol. Rev.* **80**, 555–592 (2002).
5. del Camino, D. & Yellen, G. *Neuron* **32**, 649–656 (2002).
6. Noda, M. et al. *Nature* **312**, 121–127 (1984).
7. Gandhi, C. S. & Isacoff, E. Y. *J. Gen. Physiol.* **120**, 455–463 (2002).
8. Ruta, V. et al. *Nature* **422**, 180–185 (2003).
9. Yang, N., George, A. L. & Horn, R. *Neuron* **16**, 113–122 (1996).

Optics

Positively negative

John Pendry

An artificially created material with negative refractive index has opened the door to new phenomena — and controversy. New work finally sets the seal of experimental confirmation on negative refraction.

As light crosses a boundary between different materials (such as between air and water), its speed changes and it refracts, or bends. On entering most materials, light refracts with a positive angle, as shown in Fig. 1. But some years ago, Veselago¹ suggested that some materials might produce 'negative refraction': light would refract the other way, through a negative angle. In 2001, the observation of negative refraction was reported in an artificially created material² — but not everyone was convinced. Now two papers in *Physical Review Letters*, one by Parazzoli et al.³, the other by Houck et al.⁴, report experiments at radio frequencies that confirm the existence of negative refraction.

The response of a material to electric and magnetic fields is characterized by its permittivity, ϵ , and its permeability, μ , respectively. Veselago argued that if both ϵ and μ were negative, it follows that the refractive index of the material, which determines the velocity of light within it, would also be negative. There matters rested, because, although there are natural materials with negative ϵ , there are none with both negative ϵ and negative μ . In the absence of any materials on which to experiment, the concept was purely theoretical.

But the situation changed with the publi-

cation of three papers. In the first, my own group showed⁵ that it is easy to fabricate an artificial material with negative ϵ using a lattice of thin metal wires. We then showed⁶ how to do the same thing for μ : the required magnetic response was obtained from a lattice of metal 'split rings' that resonate with magnetic fields of a given frequency; the induced currents give a negative magnetic response. The third paper, by Smith and colleagues⁷, made the key advance. This team made a structure that combined these elements in a microwave experiment, the first demonstration of negative ϵ and negative μ in the same material. They went on to demonstrate negative refraction in their material².

This work sparked great interest in negative materials, but also a fair bit of controversy: one group of sceptics wrote⁸ of "Wave Refraction in Negative-Index Media: Always Positive...". Others claimed, unfairly in my view, that the results were an artefact of losses in the system and observations taken too close to the sample. The challenge to experimenters was to reproduce the results and eliminate all areas of doubt.

Parazzoli et al.³ and Houck et al.⁴ follow the same basic design for the negatively refracting material: split rings of copper are

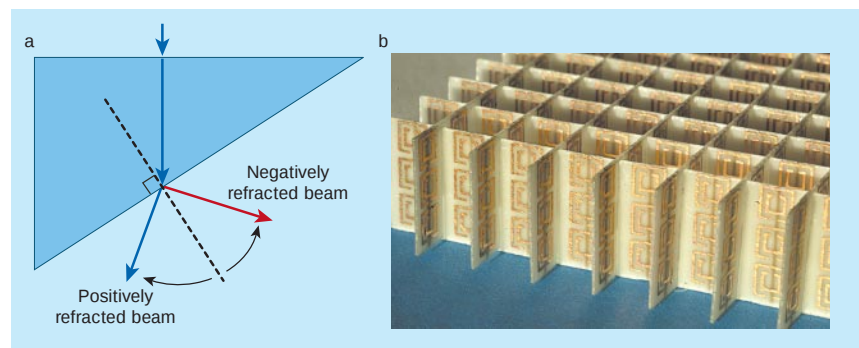


Figure 1 Negative refraction. a, Light incident on a normal material refracts at a positive angle (blue), but in a negative-index material the refraction angle is negative (red). Negative refraction occurs only in specially engineered materials. b, This arrangement of fibreglass sheets (1 cm high) in which is embedded an array of copper loops and wires was the first 'material' for which negative refraction was seen².

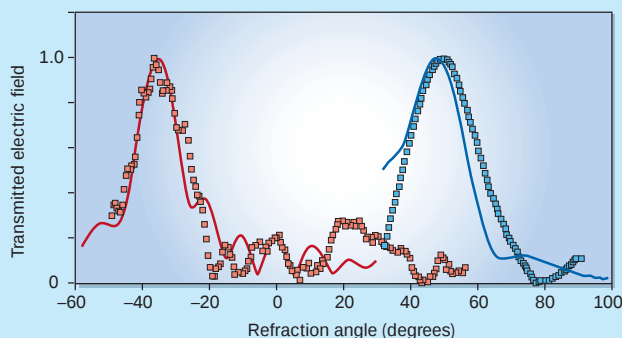


Figure 2 Negative refraction confirmed. These data from Parazzoli *et al.*³ compare refraction in their negative-index material (red points) with the response of a normal, positive-index material, Teflon (blue points). Moreover, the experimental data are in fair agreement with simulations (solid lines).

held in place by fibreglass sheets, and current circulating in them produces a negative magnetic response; in addition, thin straight copper strips respond to electric fields, also in a negative manner (Fig. 1). The new ingredients, different in each of the two experiments, are in the manner the data are collected. Parazzoli *et al.* made their measurements in free space, up to 66 cm from their material sample. The data show a clear, well-defined, negatively refracted signal (Fig. 2). Houck *et al.* work within the confines of a planar waveguide, but sample the electromagnetic field at many points. They too observe unambiguous negative refraction. In most people's minds this will settle the long-running argument about the issue of negative refraction, which has also now been lent more theoretical support by extended computations⁹.

Houck *et al.*⁴ go on to tackle another, even more controversial issue. Veselago¹ observed that a slab of negative-index material would focus light, like a lens. Because they measured the transmitted light at many points around the sample, Houck *et al.* were able to verify some limited focusing capabilities of their material. This is an area in which further work can be expected.

It has been said that tackling a new scientific problem is like going into a darkened room. First you fall over the furniture, then you collide with other people in the room; arguments might develop. With time things settle down, as you learn where most of the furniture is and don't fall over so often. Eventually someone finds the light switch and everything becomes obvious.

I think I just heard the click of a switch. ■

John Pendry is in the Condensed-Matter Theory Group, The Blackett Laboratory, Imperial College, London SW7 2AZ, UK.

e-mail: j.pendry@imperial.ac.uk

1. Veselago, V. G. *Sov. Phys. Usp.* **10**, 509–514 (1968).
2. Shelby, R. A., Smith, D. R. & Schultz, S. *Science* **292**, 77–79 (2001).
3. Parazzoli, C. G., Greger, R. B., Li, K., Koltenbah, B. E. C. & Tanielian, M. *Phys. Rev. Lett.* **90**, 107401 (2003).
4. Houck, A. A., Brock, J. B. & Chuang, I. L. *Phys. Rev. Lett.* **90**, 137401 (2003).
5. Pendry, J. B., Holden, A. J., Stewart, W. J. & Youngs, I. *Phys. Rev. Lett.* **76**, 4773–4776 (1996).
6. Pendry, J. B., Holden, A. J., Robbins, D. J. & Stewart, W. J. *IEEE Trans. Microw. Theory Techniques* **47**, 2075–2084 (1999).
7. Smith, D. R., Padilla, W. J., Vier, D. C., Nemat-Nasser, S. C. & Schultz, S. *Phys. Rev. Lett.* **84**, 4184–4187 (2000).
8. Valanju, P. M., Walser, R. M. & Valanju, A. P. *Phys. Rev. Lett.* **88**, 187401 (2002).
9. Foteinopoulou, S., Economou, E. N. & Soukoulis, C. M. *Phys. Rev. Lett.* **90**, 107402 (2003).

Genomics

Relative pathogenic values

Julian Parkhill and Colin Berry

The bacterium that causes anthrax has several close relatives. Comparison of their genome sequences should provide insight into the biology of these organisms as agents of disease — and of terrorism.

Particular notoriety has been accorded to *Bacillus anthracis* of late. As the causative agent of anthrax, this bacterium was used in the 2001 postal attacks in the United States¹, and it has reportedly been 'weaponized' as a warfare agent on at least one occasion². On pages 81 and 87 of this issue, Read *et al.*³ and Ivanova *et al.*⁴

bring the power of genomics to bear on efforts to understand this pathogen and its close relatives.

Bacillus anthracis is a member of a group of closely related organisms that includes *B. cereus*, an opportunistic pathogen of humans, and *B. thuringiensis*, an insect pathogen that has been used worldwide as



100 YEARS AGO

The discovery by Monsieur and Madame Curie that a sample of radium gives out sufficient energy to melt half its weight of ice per hour has attracted attention to the question of the source from which the radium derives the energy necessary to maintain the radiation; this problem has been before us ever since the original discovery by Becquerel of the radiation from uranium. It has been suggested that the radium derives its energy from the air surrounding it, that the atoms of radium possess the faculty of abstracting the kinetic energy from the more rapidly moving air-molecules while they are able to retain their own energy when in collision with the slowly moving molecules of air. I cannot see, however, that even the possession of this property would explain the behaviour of radium... I think that the absence of change in the radium has been assumed without sufficient justification; all that the experiments justify us in concluding is that the rate of change is not sufficiently rapid to be appreciable in a few months. There is, on the other hand, very strong evidence that the substances actually engaged in emitting these radiations can only keep up the process for a short time; then they die out... I have recently found that water from deep wells in Cambridge contains a radio-active gas, and I am anxious to see whether water from other sources possesses the same property. I should be greatly obliged if any of your readers who have access to deep level water would fill a clean two-gallon can with it and forward it to the Cavendish Laboratory. I should, of course, pay the carriage and return the can. J. J. Thomson
From *Nature* 30 April 1903.

50 YEARS AGO

Royal Society: Foreign Members. The following have been elected foreign members of the Royal Society: Louis Victor Pierre Raymond, Prince de Broglie (Paris), distinguished for his contributions to quantum theory; Marie Jules Constant Robert Courrier (Paris), distinguished for his contributions to endocrinology; Hermann Joseph Muller (Bloomington, Indiana), distinguished for his contributions to the chromosome theory of heredity; Wolfgang Pauli (Zurich), distinguished for his contributions to theoretical physics, in particular the formulation of the exclusion principle.
From *Nature* 2 May 1953.

a pesticide for over 40 years. Read *et al.*³ describe the genome sequence of a strain of *B. anthracis* that was isolated from a Texan cow and subsequently used as a standard laboratory strain around the world. The same group has previously shown that this strain is virtually identical to that used in the US postal attacks⁵ and is, therefore, still representative of a fully virulent strain. Read *et al.*³ also generated an incomplete sequence of a strain of *B. cereus* for comparative purposes, and manufactured a microarray to extend this comparison to further strains of *B. cereus*. Ivanova *et al.*⁴ have completed the sequence of a different strain of *B. cereus*, and compared it to the unfinished sequence of the postal bioterror strain generated by Read *et al.* in their earlier work.

To understand the biology of this group of organisms, one must first realize that bacteria often carry more than one form of genetic material. In addition to the chromosome that bears the bulk of the genes, bacteria often carry extra DNA molecules called plasmids — small, self-replicating and often mobile elements that carry genes for accessory functions. Plasmid functions are extremely diverse, ranging from the undetectable to the essential — the genes concerned often encode 'virulence factors'; proteins that confer resistance to antibiotics; and, crucially in this context, toxins.

It has been known for some time that plasmids confer distinct capabilities on the various members of the *B. cereus* group — which include *B. cereus* itself, *B. anthracis*, *B. thuringiensis* and related bacteria. For example, the plasmids of *B. anthracis*, pXO1 and pXO2, encode toxin and other proteins

essential for full virulence in animals, and a range of *B. thuringiensis* plasmids carry insect toxins that allow the bacterium to attack its favoured hosts. In addition, the size of the *B. cereus* genome is very variable between isolates, ranging from about 2.4 megabases to 6.4 megabases⁶. In contrast, *B. anthracis* isolates represent a group that is closely related genetically, probably indicating a recent evolutionary origin⁷. The unanswered questions around these certainties concern exactly how closely these organisms are related. Are they really separate species or, as has been suggested, the same species carrying different plasmids⁸?

The new work^{3,4} goes a long way towards answering these questions. It is clear from the results that the genomes sequenced are highly conserved, and that even many of the *B. anthracis* plasmid genes are present in other strains of the *B. cereus* group. The only significant and consistent differences between *B. anthracis* and *B. cereus* are the genes encoding toxin components, which are carried on an apparently recent insert (a pathogenicity island) within the *B. anthracis* pXO1 plasmid, and a loss by mutation in the *B. anthracis* genome of a regulator (PlcR) of the expression of a wide range of genes, many associated with virulence⁹. It is also clear from these analyses that factors involved in virulence are not restricted to the plasmid. Instead, the chromosome itself, which is often considered to be a relatively inert background for the virulence-determining plasmids, carries many genes that seem to be involved in virulence.

Read *et al.*³ and Ivanova *et al.*⁴ found that genes involved in pathogenicity for animal

and insect hosts were present on the chromosomes of both *B. cereus* and *B. anthracis*, and in many other representatives of the species. These include genes similar to those known to be involved in the pathogenicity of *Listeria* (a common food-poisoning organism in humans), and genes similar to those used by insect pathogens to attack the gut walls of their hosts. Far from being a repository of the staid metabolic genes, the chromosome is replete with its own arsenal of pathogenicity determinants. This suggests that the individual plasmids are not the only controls of virulence, but are instead more like accessory factors that can drive the overall virulence capabilities of the organism towards individual specific hosts.

This conclusion has led both groups^{3,4} to reassess the potential evolutionary niche of the ancestral organism of the *B. cereus* group. Using the genome to probe the potential metabolism of the organisms, they show that, unlike many of their relatives found in the wider environment, these bacteria seem to shun simple sugars and other carbohydrates of plant origin as a source of nutrients. Rather, they favour a more carnivorous diet of proteins, amino acids and the complex carbohydrates used by insects to build their bodies — a dietary foible taken to further extremes by a more distantly related insect pathogen, *B. sphaericus*, which is completely unable to metabolize any carbohydrates. All of this suggests that the immediate ancestor of the *B. cereus* group species was not simply a harmless soil-dwelling organism, but that it may have been saprophytic or even parasitic — that is, have preyed on the dead, or living, bodies of

Earth science

Subduction the hard way

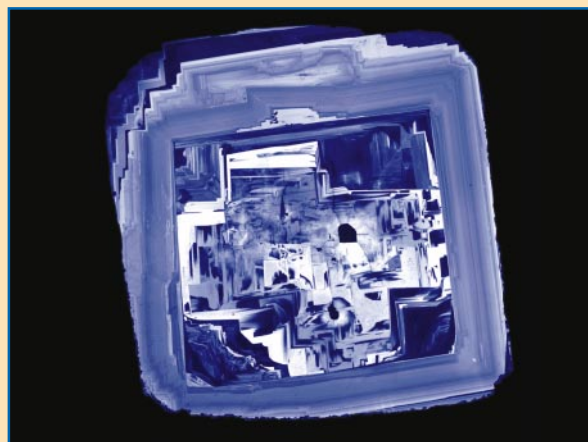
Diamonds are a geologist's best friend. Trapped inside the hardest natural substance is a relatively pristine record of the history of the continents — a record that can be exploited to investigate whether material that once formed the ocean floor of the early Earth was subducted into the mantle and preserved beneath the continental landmasses.

Earlier support for this view came from mantle rocks known as eclogite xenoliths. These are fragments of green, basalt-like material that were transported to the surface in volcanic magma. The ratio of the oxygen isotopes ¹⁸O and ¹⁶O in the eclogites suggested that they had originated as altered basalt on the ocean floor.

But the hot, turbulent past of the xenoliths means that there is no guarantee that this geological record is pure.

In this issue (*Nature* **423**, 68–70; 2003), Daniel Schulze and colleagues report their analysis of diamonds mined in Guaniamo, Venezuela. This image, captured through cathodoluminescence, shows one of their samples, 2 mm across and known as 'Picasso's diamond' for its resemblance to the cubist masterpieces of the Spanish painter.

Trapped inside this diamond (and others like it) is coesite, a form of silicon dioxide. The oxygen-isotope ratio in the coesite matches that of altered ocean-floor



basalt — "compelling evidence", say Schulze *et al.*, in favour of subduction of oceanic plates

being instrumental in the formation of the early continents.

Alison Wright

NICOLA CAVIER, ALISON DIAS

insects and other animals. Indeed, organisms that seem to be members of the *B. cereus* group have been isolated from insect digestive tracts and have been proposed to be gut-attached stages of the bacterium¹⁰. This predisposition of the *B. cereus* group may have been built upon and fine-tuned by the ready exchange of plasmids encoding accessory factors that allow the bacteria to exploit other animals or insects more directly as hosts and sources of nutrients.

All of this serves to underline the point that, whatever crude attempts are made by human beings, the true biowarfare experts are the bacteria themselves — they are constantly ready and exquisitely able to adapt to, and exploit, any environmental or pathogenic niche that presents itself. We can only hope that the availability of these sequences

will assist those who are working to protect us against these agents of disease. ■

Julian Parkhill is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

e-mail: parkhill@sanger.ac.uk

Colin Berry is at the Cardiff School of Biosciences, Cardiff University, Museum Avenue, Cardiff CF10 3US, UK.

e-mail: berry@cf.ac.uk

1. Jernigan, D. B. *et al.* *Emerg. Infect. Dis.* **8**, 1019–1028 (2002).
2. Seelos, C. *Nature* **398**, 187–188 (1999).
3. Read, T. D. *et al.* *Nature* **423**, 81–86 (2003).
4. Ivanova, N. *et al.* *Nature* **423**, 87–91 (2003).
5. Read, T. D. *et al.* *Science* **296**, 2028–2033 (2002).
6. Carlson, C. R. & Kolsto, A.-B. *Mol. Microbiol.* **13**, 161–169 (1994).
7. Keim, P. *et al.* *J. Bacteriol.* **179**, 818–824 (1997).
8. Helgason, E. *et al.* *J. Bacteriol.* **66**, 2627–2630 (2000).
9. Mignot, T. *et al.* *Mol. Microbiol.* **42**, 1189–1198 (2001).
10. Margulis, L. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 1236–1241 (1998).

Chemical physics

How to keep dry in water

Philip Ball

What does water look like close to biological surfaces? The question has provoked heated debate for decades. Experiments suggest that this 'vicinal' water may be markedly different from the bulk liquid.

Water is notorious for behaving strangely, but it is possible that we don't yet know the half of it. Amid the crush of biological macromolecules and membranes in a cell, the water of the cytoplasm rarely achieves thicknesses of more than a few molecular layers. Might this confinement between macromolecular surfaces induce behaviour very different from that of the bulk liquid? A flurry of recent papers^{1–3}, in *Langmuir* and *Physical Review Letters*, lends support to the idea — but there is, as yet, no consensus in sight⁴.

Liquid water is held together by a random, fluctuating, three-dimensional network of hydrogen bonds. This unique liquid-state structure is responsible for many of water's anomalies, and it has long been thought to play a role in the hydrophobic interaction. This attraction between hydrophobic surfaces in water is one of the key stabilizing forces of protein folding and of multi-subunit protein assemblies.

Close to a hydrophobic surface the hydrogen-bonding pattern of water is disrupted, without the compensatory interactions that might operate at hydrophilic surfaces. How does water cope with this loss of stabilization? The classic model of hydrophobic interactions⁵ proposes that, around hydrophobes, water molecules lock into a more rigidly defined cage that preserves hydrogen bonding in an 'ice-like' solvation shell. The attraction then results from the gain in entropy as some of this solid-like water is liberated by the overlapping of solvation shells when the

two hydrophobes come together. It is an appealing picture, but turns out not to be supported by any evidence of increased ordering of water around small hydrophobic species⁶.

Modern theories of inhomogeneous fluids show that in fact one needn't invoke water's unusual structure to anticipate that strange things will happen at interfaces. Capillary condensation in narrow pores shows how confinement shifts the phase diagram of a fluid. Conversely, when the interactions between a liquid and a surface are relatively unfavourable (as with water and hydrophobes), there may be a lower fluid density at the interface, leading in the extreme case to the formation of a gas-like layer (a process known as 'drying').

Lum *et al.*⁷ predicted that this is just what should happen between hydrophobic objects of various shapes (plates and cylinders, for example, approximating protein surfaces): if the surfaces are large enough and close enough together, capillary evaporation creates a bubble between them. This theoretical model included none of the geometrical nuances of a hydrogen-bonded network; it was generic to any confined fluid.

Yet water will not give up its special status so readily. There is, for example, the mystery of the long-range hydrophobic attraction. Distinct from the shorter-range force that binds proteins, this interaction between hydrophobic surfaces seems to extend over distances of up to 100 nm — equivalent to many hundreds of molecular diameters. It

has compelled some researchers to seek explanations, seemingly fantastic at face value, in terms of some kind of extended, collective ordering of the intervening water molecules.

The long-range hydrophobic interaction was potentially unified with the issue of drying at a single interface by the suggestion of Attard and co-workers⁸ that the force might arise from the growth and bridging of sub-microscopic bubbles between the surfaces. The bridging meniscus would pull the surfaces together. Tyrrell and Attard reported⁹ that, using an atomic force microscope (AFM), they had observed pancake-shaped 'nanobubbles' on hydrophobic surfaces.

Might the formation of these bubbles be promoted by partial drying at the interface — that is, by depletion of water density in this region? At least two of the new studies support this notion. Noting that in the earlier work the AFM itself might have nucleated the bubbles, Steitz *et al.*¹ have used a less invasive technique — neutron reflectivity — to look at the interface. Neutrons are strongly scattered by deuterium (D), and the researchers look at 'heavy' water, D₂O, in contact with deuterated polystyrene, a hydrophobic polymer. The scattering densities of these two substances are nearly equal, so any inhomogeneities at the interface should show up in an otherwise 'uniform' sample.

Steitz *et al.* see a surface layer 2–5 nm thick with a density about 6–12% lower than that of bulk D₂O, which they interpret as evidence of partial drying. (It is hard, however, to rule out the possibility of an apparent density deficit arising from surface migration of protonated impurities, which scatter neutrons more weakly.) They also use an AFM to reveal a surface covered with flat bubbles, each 50–120 nm wide and up to 18 nm high. They propose that the thin depletion zone acts as a precursor to the nanobubbles (which might still be nucleated by the AFM tip itself).

Jensen *et al.*² find a similar depletion zone in X-ray reflectivity measurements of the interface of a heavy-alkane monolayer floating on the surface of water. Again, the density depletion is around 10% of the bulk water density, corresponding to about one H₂O molecule for every 0.25–0.30 nm² of surface. But the depletion layer appears to be thin — less than 1.5 nm, perhaps reduced by 'capillary waves' that ruffle the surface of the free monolayer.

Another neutron-reflectivity study has been conducted by Schwendel *et al.*³, who confess that their studies on contrast-matched mixtures of D₂O–H₂O against alkylated surfaces show a density deficit in the first 2 nm or so that is so large as to be unphysical. They suspect that nanobubbles or other air inclusions might be distorting the results. In a personal communication, Thomas and co-workers indicate that they

have also found a water depletion layer at a hydrophobic surface using neutron scattering, extending over a thickness of about 1.8 nm. They are confident that their surface is free of nanobubbles, and fairly sure that protonated impurities are not responsible for the result.

These studies are starting to produce a consistent picture of hydration at hydrophobic interfaces: a thin layer of low-density, somewhat gas-like water that precedes complete capillary evaporation between two such surfaces, and helps to nucleate nanobubbles that create an apparently long-range attraction. But that may not be the end of the story. For example, Yaminsky and Ohnishi⁴ argue that experimental imperfections (protrusions from the surfaces), rather than nanobubbles, create the long-range hydrophobic force, and claim that their surface-force measurements show nothing inconsistent with the standard theory of colloidal interactions — that is, nothing unique to water — down to separations of 3 nm or so (at which point the attraction between the surfaces is so great that they spring into contact).

Moreover, would we necessarily expect 'depleted' water to be really more gas-like, or instead more ice-like, with enhanced ordering, as originally supposed⁵? Evidence of ice-like water inside a (hydrophobic) carbon nanotube at room temperature has recently been seen in simulations¹⁰. And what are the consequences for the hydration environment of proteins, where, for example, a local change in water density might be expected to affect important quantities such as pH

and salt concentration? In other words, what does the physics of confined water have to say about cell biology?

Philip Ball is a consultant editor for Nature.

1. Steitz, R. *et al.* *Langmuir* **19**, 2409–2418 (2003).
2. Jensen, T. R. *et al.* *Phys. Rev. Lett.* **90**, 086101 (2003).
3. Schwendel, D. *et al.* *Langmuir* **19**, 2284–2293 (2003).
4. Yaminsky, V. & Ohnishi, S. *Langmuir* **19**, 1970–1976 (2003).

5. Frank, H. S. & Evans, M. W. *J. Chem. Phys.* **13**, 507–532 (1945).
6. Blokzijl, W. & Engberts, J. B. F. *N. Angew. Chem. Int. Edn* **32**, 1545–1579 (1993).
7. Lum, K., Chandler, D. & Weeks, J. D. *J. Phys. Chem. B* **103**, 4570–4577 (1999).
8. Parker, J. L., Claesson, P. M. & Attard, P. *J. Phys. Chem.* **98**, 8468 (1994).
9. Tyrrell, J. W. G. & Attard, P. *Phys. Rev. Lett.* **87**, 176104 (2001).
10. Mashl, R. J. *et al.* *Nano Lett.* doi:10.1021/nl0340226 (2003).

Molecular biology

Complicity of gene and pseudogene

Jeannie T. Lee

'Pseudogenes' are produced from functional genes during evolution, and are thought to be simply molecular fossils. The unexpected discovery of a biological function for one pseudogene challenges that popular belief.

Pseudogenes are defective copies of functional genes that have accumulated to an impressive number during mammalian evolution¹. Dysfunctional in the sense that they cannot be used as a template for producing a protein, pseudogenes are in fact nearly as abundant as functional genes^{2,3}. Why have mammals allowed their accumulation on so large a scale? One proposed answer is that, although pseudogenes are often cast as evolutionary relics and a nuisance to genomic analysis, the processes by which they arise are needed to create whole gene families⁴, such as those involved in immunity and smell. But are pseudogenes themselves merely by-products of this process? Or do the apparent evolutionary

pressures to retain them hint at some hidden biological function? For one particular pseudogene, the latter seems to be true: elsewhere in this issue (page 91), Hirotsune and colleagues⁵ report the unprecedented finding that the *Makorin1-p1* pseudogene performs a specific biological task.

Hirotsune *et al.*⁵ had been analysing mice in which copies of a fruitfly gene called *Sex-lethal* were randomly inserted in the mouse genome. In the course of their studies, they encountered one mouse line that died shortly after birth from multi-organ failure. As this occurred in only one mouse line out of many, the results could not be explained by aberrant *Sex-lethal* expression. Instead, the authors attributed their finding

Genetics

Suicidal mushroom cells

Programmed cell death — apoptosis — is a universal phenomenon among multicellular organisms, and is especially important during development. Genetically orchestrated mechanisms of cell death have also been found in single-celled protists and yeast. Writing in *Fungal Genetics and Biology* (39, 82–93; 2003. doi: 10.1016/S1087-1845(03)00024-0), Benjamin Lu and colleagues describe a remarkably simple version of apoptosis in the ink-cap mushroom *Coprinus cinereus* (pictured).

Lu *et al.* studied apoptosis in mutant strains of the fungus that have defects in spore formation. Spores are formed by meiosis, which is the same type of cell division that halves the number of chromosomes in the egg and sperm cells of animals. In the mushroom, meiosis occurs in a synchronous fashion,



sweeping across the gill surfaces underneath the cap, reconfiguring and sorting the chromosomes within 10 million spore-producing cells called basidia (inset). Mutants of *C. cinereus* called 'white-caps' are infertile: their basidia show defects at various points in the cell

cycle, and the ashen hue of their delicate umbrellas is caused by the resulting failure to form black-pigmented spores.

It turns out that, in the mutants, basidia that experience problems at the beginning of meiosis (prophase I) undergo mass apoptosis, showing the classical apoptotic hallmark of DNA fragmentation. Lu and colleagues' experiments suggest that apoptosis is triggered at a single checkpoint in the mushroom cell cycle. This contrasts with the situation in the mouse, where the switch that activates apoptosis can be tripped at many steps throughout meiosis, implying that there is an almost continuous molecular assessment of the viability of the gamete-forming cells. Not surprisingly, it seems that

mushrooms lack some of the developmental sophistication of mammals.

But why should apoptosis occur in mushrooms at all? Resource conservation is a likely explanation. Mushrooms disperse astonishing numbers of airborne spores, but there is an infinitesimal chance that any individual spore will reach a suitable patch of ground, survive, out-compete the resident microbes, grow and one day find a mate. So by aborting any basidia that have mishaps early in meiosis, the fungus conserves resources for healthy cells that will succeed in producing viable spores. **Nicholas P. Money**
Nicholas P. Money is in the Department of Botany, Miami University, Oxford, Ohio 45056, USA. e-mail: moneynp@muohio.edu

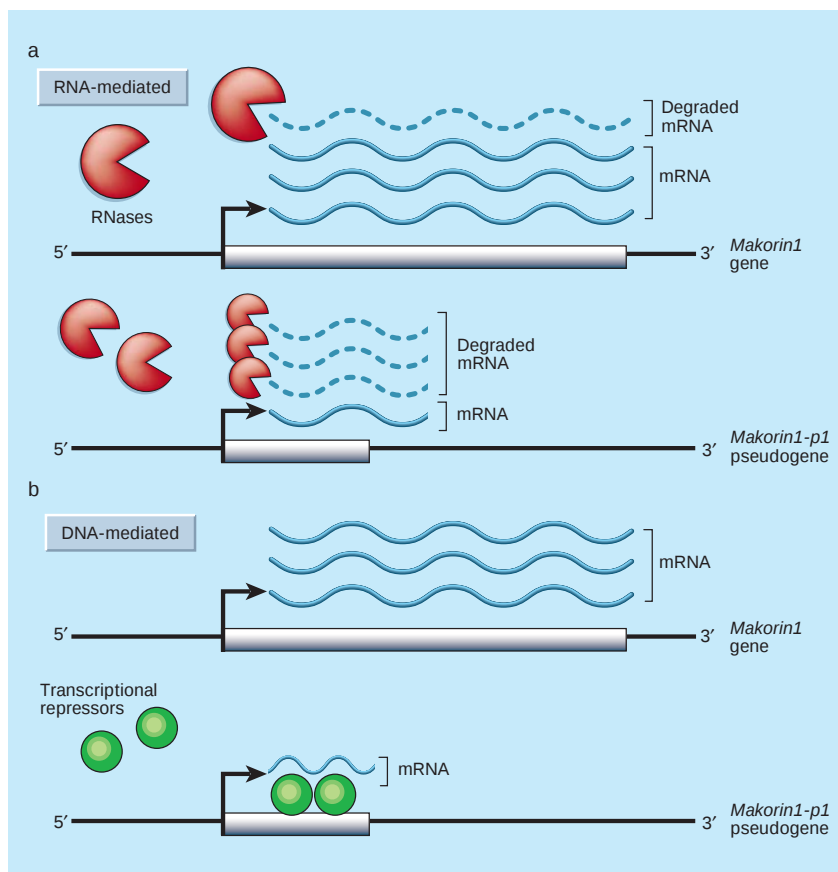


Figure 1 Gene regulation by a pseudogene. Hirotune *et al.*⁵ have found that the expression of the *Makorin1* gene is controlled by one of its pseudogene copies, *Makorin1-p1*. The figure shows two ways in which this might happen. **a**, An RNA-mediated mechanism. Here, messenger RNA copies of the pseudogene and gene compete for a destabilizing protein that binds a crucial 700-nucleotide region near the beginning of the mRNAs. This destabilizing protein might be an RNA-digesting enzyme (RNase). **b**, A DNA-mediated mechanism. Here, regulatory elements in the 700-nucleotide region of the pseudogene and gene compete for transcriptional repressors.

to a disruption of the particular stretch of genomic information into which *Sex-lethal* had inserted in this case. Whereas some might have dismissed the line as an aberration and unworthy of the effort required to characterize it, Hirotune and colleagues delved deeper and were rewarded with the surprising finding that a pseudogene can regulate the expression of the functional gene from which it arose.

The authors first found that, in the mouse line in question, the inserted *Sex-lethal* gene disrupted *Makorin1-p1* — a pseudogene copy of the functional *Makorin1* gene. Only recently identified⁶, *Makorin1* is an ancient gene that has been evolutionarily conserved from nematode worms to fruitflies and mammals, and encodes a putative RNA-binding protein. It is the prototype of a large family of *Makorin* genes and pseudogenes, and is located on mouse chromosome 6.

By contrast, Hirotune *et al.* found that the pseudogene *Makorin1-p1* lies on chromosome 5. Like the original gene, this pseudogene can be 'transcribed' into a messenger RNA copy. But it has incurred

mutations during evolution, so the mRNA cannot, as is usual, be used to produce a protein. Further differences between the gene and pseudogene include the fact that the *Makorin1-p1* mRNA contains only the first (that is, 5') 700 nucleotides of the *Makorin1* mRNA. Moreover, whereas both copies (one from the mother and one from the father) of the *Makorin1* gene can be transcribed, the *Makorin1-p1* pseudogene is paternally 'imprinted', so that only the paternal copy is expressed.

Normally, *Makorin1* mRNA is expressed throughout the animal⁶. But Hirotune *et al.* found that when the paternal *Makorin1-p1* pseudogene was disrupted, the expression of *Makorin1* was markedly reduced in embryos and throughout birth and weaning. This implies that the pseudogene is normally required for the high-level expression of *Makorin1*. Interestingly, of the two forms of *Makorin1* mRNA, only the smaller 1.7-kilobase transcript was downregulated — the larger 2.9-kilobase copy was unaffected. The long and short forms are identical except in a region at the so-called 3' end

that is not translated into protein. So, it seems that this region in the long form functions independently to keep expression levels high.

The authors also wondered whether the imprinting of *Makorin1-p1* is mechanistically central to *Makorin1* expression. However, its disruption had equal effects on both maternal and paternal *Makorin1* genes. So it seems that the imprinting of *Makorin1-p1* is an odd happenstance that has little or nothing to do with its function. Rather, it seems likely that, when *Makorin1-p1* arose, it fortuitously integrated into a chromosomal region that was already imprinted.

This study⁵ generates many new and exciting questions. For instance, is *Makorin1-p1* the only *Makorin* pseudogene that regulates *Makorin1*? Considering that disruption of *Makorin1-p1* causes only a partial loss of expression of the functional gene, one might speculate that there are indeed other pseudogenes whose functions partly overlap, and that the deployment of an entire pseudogene battalion might be a feasible strategy of gene regulation.

Furthermore, how does *Makorin1-p1* regulate *Makorin1*? The authors found that the 700-nucleotide 5' region of *Makorin1-p1* not only was required but was also sufficient for regulation in experiments *in vitro*. These experiments also suggested that the pseudogene acts sequence-specifically, affecting only those genes that show some sequence similarity to itself.

Non-protein-coding RNAs have recently been shown to perform a variety of tasks, such as gene silencing, catalysis and the regulation of development⁷. So *Makorin1-p1*'s mechanism of action might involve its non-coding RNA product, rather than the pseudogene itself. Hirotune *et al.* propose that this product works to stabilize the *Makorin1* mRNA (Fig. 1a). They favour a model in which the first 700 nucleotides of the *Makorin1* mRNA contain a recognition site for a destabilization factor. Because this 700-nucleotide domain is shared by the *Makorin1-p1* mRNA, the expression of the pseudogene would provide a means of titrating out the destabilizing factor through direct competition. In this model, the longer *Makorin1* mRNA is unaffected because its 3' untranslated region protects it from degradation. An extension of this idea is that *Makorin1* self-regulates: it has been suggested that it encodes an RNA-binding protein⁶, which might be the destabilizing factor that downregulates the short form of its own mRNA.

Given the available data, however, another mechanism could be at work (Fig. 1b). This is suggested by the fact that mRNA stability is usually controlled by elements in the 3' untranslated region⁸ — rather than at the 5' end, where the key 700-nucleotide region of *Makorin1* is found. The alternative

mechanism would involve the pseudogene DNA locus directly. For example, perhaps the 700-nucleotide region in the gene and pseudogene contains elements that, on binding certain proteins, repress transcription. In this model the repressor proteins would be limited in availability, so that *Makorin1-p1* would compete for repressor binding. These two models — RNA-mediated versus DNA-mediated — have mechanistic differences and could be tested.

Whatever the underlying mechanism, the work of Hirotsune *et al.*⁵ is provocative for revealing the first biological function of any pseudogene. It challenges the popular belief that pseudogenes are simply molecular fossils — the evidence of Mother Nature's experiments gone awry. Indeed, it suggests that evolutionary forces can work in both directions. The forward direction is driven by pressures to create new genes from existing ones, an imperfect process that often generates defective copies of the original. But these defective copies need not be evolutionary dead ends, because pressures

in the reverse direction could modify them for specific tasks. In the case of *Makorin1* and *Makorin1-p1*, the result of bidirectional selection is that one gene cannot exist without the other — an example of functional complicity between a perfected product of evolution and its derivative castaway. Might the pseudogene copies of other functional genes be similarly useful? ■

Jeannie T. Lee is at the Howard Hughes Medical Institute, Department of Molecular Biology, Massachusetts General Hospital, and the Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA.
e-mail: lee@molbio.mgh.harvard.edu

1. Harrison, P. M. *et al.* *Genome Res.* **12**, 272–280 (2002).
2. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
3. Mouse Genome Sequencing Consortium *Nature* **420**, 520–562 (2002).
4. Harrison, P. M. & Gerstein, M. J. *Mol. Biol.* **318**, 1155–1174 (2002).
5. Hirotsune, S. *et al.* *Nature* **423**, 91–96 (2003).
6. Gray, T. A. *et al.* *Genomics* **66**, 76–86 (2000).
7. Eddy, S. R. *Nature Rev. Genet.* **2**, 919–929 (2001).
8. Mignone, F., Gissi, C., Liuni, S. & Pesole, G. *Genome Biol.* **3**, 4.1–4.10 (2002).

Atmospheric chemistry

Burning domestic issues

Joel S. Levine

In the developing world much of the energy for heating, lighting and cooking comes from burning 'biomass', mainly wood. A first attempt has been made to quantify the resulting emissions to the atmosphere.

Almost a quarter-century ago, Paul Crutzen and colleagues¹ published pioneering work showing how the burning of biomass produces emission of a whole variety of trace gases. Since then there has been a growing realization of the environmental significance of this source of gases, and of associated particulate material, and the effects range from the local through the regional to the global. Biomass burning also influences the biogeochemical cycling of carbon and nitrogen compounds from the soil to the atmosphere.

Although most such burning occurs during human-initiated land clearance and changes in land use, a large component is due to the use of biomass fuels for domestic activities. That component has not previously been quantified. As they describe in the *Journal of Atmospheric Chemistry*, however, Ludwig and colleagues² have now provided some estimates, and the figures concerned are substantial.

Trace gases produced during biomass burning and released into the atmosphere include carbon dioxide (CO₂), carbon monoxide (CO), methane (CH₄) and non-methane hydrocarbons, hydrogen (H₂), nitric oxide (NO), ammonia (NH₃), methyl chloride (CH₃Cl) and sulphur species³.

Carbon dioxide and methane are greenhouse gases that lead to global warming. Nitric oxide and sulphur species lead to the photochemical production of nitric acid and sulphuric acid, two of the main components



Figure 1 Domestic duty — an Ethiopian villager with firewood collected for cooking.

of acid precipitation. Methane, carbon monoxide and nitric oxide bring about the photochemical production of ozone in the troposphere. Ozone is a pollutant and irritant, as well as a greenhouse gas.

In studies of this topic, it is usual to divide burning into the following components: forests (tropical, temperate and boreal); savannas; agricultural waste left after harvesting (cereals' stubble, for instance); and fuel wood for domestic heating and cooking. For the first three components, the geographical distribution and area burned are key parameters in calculating the amount of gases and particulates released into the atmosphere. Although there has never been a dedicated satellite to monitor and quantify burning, satellite instrumentation designed for other purposes has allowed identification of the location of active fires and burned scar areas in forests, savannas and agricultural regions⁴. Space-based measurements can provide little information on the fourth component. But Ludwig *et al.*² point out that the use of fuel wood, charcoal and non-woody biofuels for cooking, heating and lighting is a daily event for about half of the world's population, mostly in the developing world (Fig. 1), and they have produced some detailed estimates of the resulting emissions.

That task required two types of information: data on the consumption of biomass fuels, and data on the emission of gases and particulates per unit quantity of those fuels (the emission factors). Ludwig *et al.* collected the limited published data on the first topic, and in assessing the second they performed a series of laboratory measurements of the emission factors of domestic biomass burning. To estimate the contribution to the global inventory of trace-gas emissions, they assumed that 80% of domestic biomass burned is wood, 15% is agricultural residue, 2.5% is dung and 2.5% is charcoal. Finally, they assumed that about 85% of the domestic emissions are taking place in the developing countries of Asia, Africa and Latin America.

As Ludwig *et al.* themselves say, the calculations have a high degree of uncertainty. Nonetheless, the figures — given here in teragrams (1 Tg is 10¹² grams, or 10⁶ tonnes) — are illuminating. The estimated annual global release through domestic biomass burning is 1,495 Tg of carbon in the form of CO₂, 141 Tg of carbon in the form of CO, and 2.54 Tg of nitrogen in the form of NO. These are significant proportions of the annual global production of these environmentally important gases: 17% of total CO₂, 13% of total CO and 6% of total NO. But as the authors point out, the estimated CO₂ release is not necessarily a net emission, as it depends on the sustainability of wood fuels. The use of agricultural residues and dung can be assumed to be 100% sustainable and

therefore contributes no CO₂ net release to the atmosphere. For the other compounds, however, the emissions are net releases. Finally, as far as checks and balances are concerned, Ludwig *et al.* note that although biomass burning for domestic purposes might well increase as the population increases, that will diminish the stock of fuel for wild fires.

This paper² constitutes a new chapter in our understanding of biomass burning as a source of environmentally important atmospheric trace gases. Ironically, atmos-

pheric chemists will be appraising the lessons to be learned from it at the same time as they assess the environmental impact of a more dramatic component of fuel combustion — the burning of Iraqi oil fields. ■

Joel S. Levine is at the NASA Langley Research Center, Hampton, Virginia 23681, USA.

e-mail: j.s.levine@larc.nasa.gov

1. Crutzen, P. J., Heidt, L. E., Krasnec, J. P., Pollock, W. H. & Seiler, W. *Nature* **282**, 253–256 (1979).
2. Ludwig, J. *et al.* *J. Atmos. Chem.* **44**, 23–37 (2003).
3. Crutzen, P. J. & Andreae, M. O. *Science* **250**, 1669–1678 (1990).
4. Levine, J. S. (ed.) *Biomass Burning and Global Change*, Vols 1 and 2 (MIT Press, Cambridge, Massachusetts, 1996).

Astronomy

Elements of surprise

John Cowan

The discovery of a very distant galaxy for which the abundances of around 25 elements can be measured promises new insight into the history of element creation and star formation in the Universe.

In the Big Bang at the beginning of the Universe, the lightest elements, hydrogen, helium and lithium, were created. Two other light elements (beryllium and boron) are produced in interstellar space in interactions between cosmic-ray particles and gas atoms¹. But all of the other elements that exist in nature have been synthesized in nuclear reactions — ‘nucleosynthesis’ — inside stars, from where they are ejected into interstellar space and eventually find their way into new stars and planets.

Astronomers have made detailed studies of the synthesis of elements in the Milky Way² and in some relatively nearby galaxies³. But little was known about the production of elements, and the associated history of star formation, in the most distant galaxies that formed early in the history of the Universe. On page 57 of this issue, Prochaska *et al.*⁴ report observations of the abundance of elements in a galaxy far away and less than 2.5 billion years old. Their work opens a new window on the early formation of elements and stars in the Universe.

A number of studies have explored ‘damped Lyman alpha’ (DLA) systems, in which clouds of hydrogen gas are detected through the radiation they absorb from even more distant quasars. These studies probe the earliest chemical history of gas in the Universe, the gas that would form the first stars in the first galaxies. Prochaska *et al.* were able to identify a DLA galaxy along the line of sight to a more distant quasar (known as FJ081240.6+320808). The distance of an object is usually indicated in terms of its ‘redshift’ — how much the wavelength of its emitted light has increased on its way to Earth, due to the expansion of the Universe. The DLA galaxy has a redshift, *z*, of 2.626 and the quasar of 2.701. Such large shifts towards

the red end of the spectrum indicate that these objects are at great distances: in this case, it took almost 12 billion years for the light from this galaxy to reach Earth.

It is significant that this ‘galaxy at redshift *z* = 2.626’ (its only designated name so far) is the first distant galaxy to be found that, because of its substantial number of sufficiently abundant elements, is suitable for additional, detailed abundance studies. Prochaska *et al.* followed up their initial discovery with high-resolution studies of the galaxy’s radiation spectrum using the HIRES spectrograph of the Keck I telescope in Hawaii. In contrast to many earlier studies that were limited only to intergalactic gas clouds and only a few elements, these authors observed approximately 25 elements in the galaxy at *z* = 2.626, including a number of heavy elements such as zinc and germanium.

The presence of these elements, particularly those heavier than iron, in such a young galaxy is striking. Fundamentally, it seems to indicate that in the galaxies (or at least in this galaxy) that formed relatively shortly after the Big Bang, the onset of star formation and related element production was very rapid. Indeed, Prochaska *et al.* argue in favour of nucleosynthesis in massive stars, which could have formed rapidly and which have very short lifetimes (a few million years), ending in violent supernova explosions. Further supporting that view is the presence of elements such as oxygen, magnesium and sulphur in this galaxy — such elements are characteristically produced in massive stars⁵.

Abundance determinations in objects as far away as the galaxy at *z* = 2.626 are complicated by the effects of dust. Some elements might be incorporated in dust particles, depleting their observed abundance in the

galactic gas. Nevertheless, based on our knowledge of this problem in our own Galaxy, it is possible to obtain total elemental abundances. Remarkably, the total abundance pattern found by Prochaska *et al.* is consistent with a scaled version of the abundance pattern for the Solar System. Because the galaxy at *z* = 2.626 was formed early in the Universe’s history and is so much older than the Solar System (which is only 4.5 billion years old), this consistency of elemental abundances suggests that there are some cosmic universalities or similarities in the synthesis history of all elements.

It should be noted, however, that only upper limits, not absolute values, were measured for the abundances of the chemically interesting heaviest elements, tin and lead, in this distant galaxy. A number of puzzles also remain in interpreting and understanding the abundance distribution. Much germanium, for example, is produced in the Solar System by a process of slow neutron-capture (known as the s-process). This nucleosynthetic process is thought to occur in low-mass stars that take billions of years to live and die⁶, and hence it could not be a means of producing germanium as early in the history of the Universe as the observations of the galaxy at *z* = 2.626 would suggest. It may be that germanium production is tied to the overall metal abundance of galactic gas, as seems to be the case in some stars in the Milky Way⁷. Also not yet quantified is the role of the other neutron-capture process (rapid or r-process), which may well be contributing to the abundances of some of these heavier elements. Further abundance studies, particularly for elements such as lead, will be needed to help untangle the history of element formation in this galaxy.

What is most encouraging about Prochaska and colleagues’ findings is that there may be other such distant and young galaxies that can be similarly studied and analysed. The authors already report a second distant galaxy, with a different redshift, but surprisingly along the same line of sight as the first galaxy that they observed. They further suggest that around 2% of high-redshift galaxies may be suitable for abundance studies. These additional probes will no doubt lead to a more complete understanding of the nature of element and star formation over the history of the Universe. ■

John Cowan is in the Department of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma 73019, USA.

e-mail: cowan@mail.nhn.ou.edu

1. Fields, B. D. & Olive, K. A. *Astrophys. J.* **516**, 797–810 (1999).
2. Sneden, C. & Cowan, J. *J. Science* **299**, 70–75 (2003).
3. Shetrone, M. *et al.* *Astron. J.* **125**, 684–706 (2003).
4. Prochaska, J., Howk, J. C. & Wolfe, A. M. *Nature* **423**, 57–59 (2003).
5. Woosley, S. E. & Weaver, T. A. *Astrophys. J. Suppl. Ser.* **101**, 181–235 (1995).
6. Busso, M., Gallino, R. & Wasserburg, G. J. *Annu. Rev. Astron. Astrophys.* **37**, 239–309 (1999).
7. Sneden, C. *et al.* *Astrophys. J.* **496**, 235–245 (1998).

Materials science

Memories made in microstructure

Nature Mater. **2**, 307–311 (2003)

Shape-memory alloys are used in a wide range of applications, from spectacle frames to instruments for keyhole surgery. Despite this technological importance, a fundamental explanation of the shape-memory effect in alloys such as NiTi ('nitinol') has so far eluded scientists. Xiangyang Huang and colleagues now suggest that shape memory might be stored in the internal stresses in the microstructure of the alloy.

Shape-memory alloys must be heated to recover their original shape after deformation from the 'parent' phase. Although each parent state can be deformed into several 'daughter' states, on heating each daughter returns to the same parent phase. How does the daughter state remember the structure of the parent phase if there is no unique crystallographic relationship between the two?

The conventional view is that there must be an atom-to-atom correspondence between the parent and daughter states. However, Huang *et al.* propose that shape memory is stored in the internal stresses. Because the stresses arise from the microstructure, they are expected to be present in both the parent and daughter phases. This also provides a simpler picture of the reverse shape-memory effect, by which nitinol can be 'trained' to remember its daughter structure — although experimental evidence will be required to confirm it.

Charlene Lobo

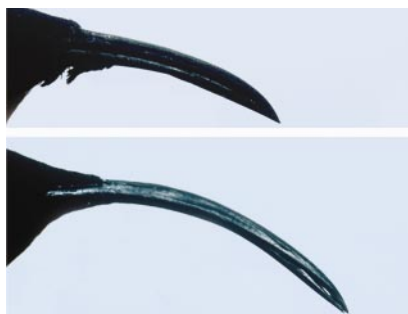
Evolution

Flowers to fit the bill

Science **300**, 630–633 (2003)

Males and females of the purple-throated carib hummingbird (*Eulampis jugularis*) are adapted to feed from different species of heliconia flower. Males have short straightish bills (pictured), designed to fit the flowers of *Heliconia caribaea*. Females are the smaller sex, but have longer, curvier bills and feed from *H. bihai*. The birds are the flowers' sole pollinators.

But it's more complicated than that, as Ethan J. Temeles and W. John Kress have discovered. On the Caribbean island of St Lucia, at sites where *H. caribaea* is rare, *H. bihai* has evolved a short and straight, male-friendly flower to fill the gap. On nearby Dominica, the two plant species are divided by altitude: *H. caribaea* in the lowlands, *H. bihai* at higher elevations. Where the two species overlap, the ecological situation is reversed, with *H. caribaea* having evolved an elongated form for females.



The short and the long of it. Male (top) and female bills of the purple-throated carib hummingbird.

As well as shape and colour, the flowers change their nectar provisioning to match their clients, with those providing for the males producing more nectar. The inter-island differences show that flower and hummingbird have apparently evolved to get the most out of their association with each other.

John Whitfield

Malaria

Pump hijacker

J. Cell Biol. **161**, 103–110 (2003)

By hijacking the membrane of red blood cells, the malaria parasite concocts for itself a bubble of calcium-rich soup, Marcos L. Gazarini and colleagues have found.

The single-celled malaria parasite, *Plasmodium falciparum*, needs concentrated calcium outside its cell so that it can let in calcium pulses that trigger various internal activities. But for part of its life cycle in humans, it dwells in red blood cells — which are low in calcium.

The parasite solves this problem by building around itself an envelope from the blood cells' plasma membrane. This previously known structure is called the parasitophorous vacuole. Gazarini *et al.* now show, using a dye that fluoresces more brightly when it binds calcium ions, that this bubble contains calcium concentrations that are 100–1,000 times higher than elsewhere in the red blood cell.

The parasite probably commandeers ion pumps already in the membrane that normally drain blood cells of calcium, using them instead to drive calcium ions into the vacuole. The authors found that blocking this ion flow prevented the parasite from maturing, suggesting that such a strategy might be used in the development of anti-malarial drugs.

Helen Pearson

Electrochemistry

Magnesium power

Adv. Mater. **15**, 627–630 (2003)

Lithium batteries have become the pre-eminent lightweight, rechargeable power source, although lead-acid and

nickel-cadmium cells are relatively cheap and still widely used in high-power applications such as electric vehicles. But cadmium and lead are toxic and heavy. In principle, magnesium-based batteries would represent a happy medium, being based on a metal that is light, cheap and environmentally friendly. Orit Chusid and colleagues have now demonstrated prototypes of viable, rechargeable magnesium batteries that withstand many recharging cycles with just a few per cent loss in power capacity.

Like lithium cells, these batteries are 'rocking-chair' devices in which magnesium ions shuttle back and forth in discharge and recharge cycles. The anode is a commercial magnesium alloy, more ductile than pure magnesium, and the cathode is Mo_6S_8 , which can intercalate magnesium reversibly. The electrolyte is a polymer gel in which magnesium ions can move easily.

The result is a battery with a discharge voltage of 0.9–1.2 V and a specific capacity that, at temperatures above 60 °C, approaches the maximum theoretical value. The authors think it should be possible to make rechargeable magnesium batteries with energy densities up to 50% greater than those of lead-acid and nickel-cadmium cells.

Philip Ball

Cell biology

Torpedo smart gel

Proc. Natl Acad. Sci. USA **100**, 3485–3490 (2003)

According to David Reigada and colleagues, synaptic vesicles from nerve cells of the electric ray *Torpedo* are filled with a matrix that acts like a smart gel. The authors propose that this provides an intelligent interface between nerves and the extracellular environment, regulating the fine-tuning of neurotransmitter release.

Neurotransmitters are the chemicals that neurons use to communicate. They are stored in synaptic vesicles, where they had been thought to exist in free solution. But Reigada *et al.* show that in the electric organ of *Torpedo*, 95% of the neurotransmitter acetylcholine (ACh) and the energy-storing molecule ATP is bound to a matrix inside the vesicles. This modulates the amount of ACh and ATP released, through an ion-exchange system. As the number of positively charged ions increases, ACh and ATP are displaced from the matrix, causing it to swell.

The matrix is a complex structure made of several proteins, but its core is formed from a vesicle membrane protein called SV2. This may be the molecule responsible for immobilizing and releasing ACh and ATP. The researchers suggest that the matrix regulates the amount of freely diffusible ACh and ATP in the early stages of neurotransmitter release.

Helen R. Pilcher

Polarized light as a butterfly mating signal

This optical feature of some iridescent wings catches a suitor's eye in the deep forest.

Iridescent butterfly scales are visually stunning structures that reflect highly saturated colour. They also create an array of non-chromatic optical phenomena, such as polarization, polarization mixing and highly directional flashes^{1,2}, but the ecological purpose of these effects is unclear^{3,4}. Here we show that polarized light is used in mate recognition by *Heliconius* butterflies, a genus that is known to rely on visual cues in sexual selection and speciation⁵. This terrestrial example of exploitation of polarized light may have adaptive value in dense forest, where illumination varies greatly in spectrum and intensity.

Coloured light from thin-film iridescence, as in the iridescent members of *Heliconius*, is often polarized¹ — for example, the blue iridescence of *H. cydno* is 90% polarized at certain angles of reflection (Fig. 1). To determine whether this polarization is involved in mate recognition, we carried out mate-choice experiments⁵ to compare responses in *H. cydno chioneus* with those of the closely related but non-iridescent *H. melpomene malleti*⁶.

We displayed female butterfly wings, which were conspecific to the males being tested, behind one of two filters: optically anisotropic, colourless, depolarizing mylar⁷ (Grafix Dura-Lar 0.003, Graphic Arts Systems) or a circular polarizing screen (3M Optical Systems Division). Control filters, chosen for their similar transmission of

ultraviolet and visible light, were polarization-neutral, isotropic colourless plastic and grey-tinted window glass, respectively. The depolarizing and circular polarizing filters remove polarization signals from wings while leaving colour signals intact; the control filters leave both polarization and colour signals unaffected.

In an outdoor insectary containing at least five conspecific males, we displayed mounted female wings beneath filters, and moved them to simulate flight. The number of male flights through a volume of radius 30 cm above the window was counted by eye over a 10-min period for the depolarized and control conditions, with the order of presentation being randomized; one pair of 10-min treatments represents a single trial. Male *H. cydno* approached the female wings significantly less often when the wings were displayed behind the depolarizing filters rather than the non-depolarizing control filters, but there was no significant difference in the response to the non-iridescent *H. melpomene* for either condition (Fig. 2).

To our knowledge, this is the first example of polarized light being used for mate recognition, or indeed for detection of any object, in a terrestrial environment. It may be particularly important in the visually complicated forest inhabited by *H. cydno* and by other iridescent *Heliconius* species, such as *H. sapho*⁸. Whereas the spectral radiance of pigmentary colour is strongly affected by the patchy light distribution in the forest, a polarized signal from iridescent wings is easily recognizable against the relatively unpolarized background^{9,10}.

The use of iridescence for intraspecific recognition may correlate with the radiation of *Heliconius* into deep-forest habitats. The occurrence of several species pairs with different iridescent properties, such as *H. melpomene* and *H. cydno*, suggests that the use of polarization has contributed to adaptive radiation and speciation.

Butterflies of several species are physiologically sensitive to polarization^{11,12}. Unlike the eyes of other insects, such as bees, which have polarization-sensitive ommatidia only in their dorsal rims, butterflies' eyes are sensitive to polarization in all of their ommatidia¹¹. Butterflies of the genus *Papilio* can also discriminate between stimuli that differ only in their angle of polarization¹². Investigation of the effects on butterfly behaviour of the optical properties and visual ecology of structural colour should help to explain patterns of ecological diversity in different butterfly groups.

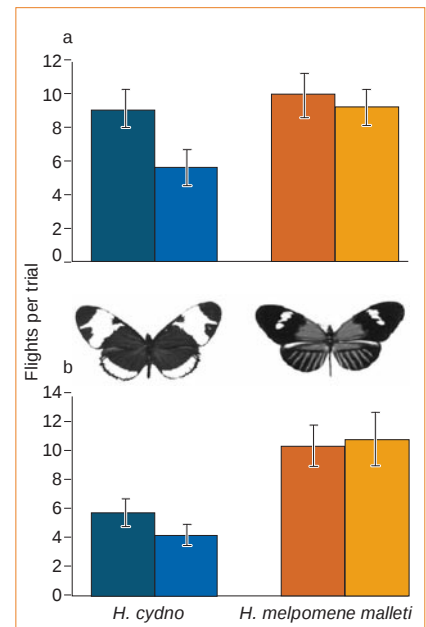


Figure 2 Response of male *Heliconius cydno* and *H. melpomene malleti* butterflies to polarized and depolarized views of female butterfly wings. **a, b**, Mean number of male approaches per 10-min trial period to a polarized (left bars) and depolarized (right bars) female stimulus. In **a**, female wings were presented through a glass window made up of either polarization-neutral grey glass (control) or a circular polarizing screen; in **b**, a plastic window consisting of polarization-neutral plastic film (control) or a mylar depolarizer was used. Error bars show standard error. For iridescent *H. cydno*, $P=0.01$ is the combined probability value of two independent experiments: $P=0.05$, $n=17$ for plastic treatment; $P=0.02$, $n=8$ for glass treatment. For the non-iridescent *H. melpomene malleti* control group, $P=0.88$ is the combined probability value of two independent experiments: $P=0.83$, $n=9$ for plastic treatment; $P=0.66$, $n=10$ for glass treatment. P values in a single experiment compare the total number of male flights to filters by Yates-corrected χ^2 test. The smaller response by *H. cydno* males in **b** is due to the time of day at which the experiments were done.

Alison Sweeney*†, Christopher Jiggins†‡, Sönke Johnsen*

*Department of Biology, Duke University, Durham, North Carolina 27708, USA

e-mail: ams27@duke.edu

†Smithsonian Tropical Research Institute, Box 2072, Balboa, Panama

‡Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK



Figure 1 Polarized iridescent patterning of the butterfly *Heliconius cydno* (top) compared with *H. melpomene malleti* (bottom), whose wings do not show polarized iridescence. Photographs of left wings are unmodified; images of right wings were generated by taking two photographs through a polarizing filter that was rotated by 90° between exposures, and then producing the difference of the two images in Photoshop (Adobe). *H. cydno* shows a pattern of polarized and depolarized regions, whereas *H. melpomene malleti* shows no polarization pattern. Wing spans, 9 cm.

1. Vukusic, P., Sambles, J. R. & Lawrence, C. R. *Nature* **404**, 457 (2000).
2. Vukusic, P., Sambles, J. R., Lawrence, C. R. & Wootton, R. J. *Proc. R. Soc. Lond. B* **269**, 7–14 (2002).
3. Kemp, D. J. *Trends Ecol. Evol.* **17**, 298–300 (2002).
4. Parker, A. *Mater. Today* **5**, 26–31 (2002).
5. Jiggins, C. D., Naisbit, R. E., Coe, R. L. & Mallet, J. *Nature* **411**, 302–305 (2001).
6. Shashar, N., Hagan, R., Boal, J. G. & Hanlon, R. T. *Vis. Res.* **40**, 457–467 (2000).

- 71–75 (2000).
 7. Beltran, M. *et al.* *Mol. Biol. Evol.* **19**, 2176–2190 (2002).
 8. Estrada, C. & Jiggins, C. D. *Ecol. Entomol.* **27**, 448–456 (2002).
 9. Endler, J. A. *Ecol. Monogr.* **63**, 1–27 (1993).
 10. Shashar, N., Cronin, T. W., Wolff, L. B. & Condon, M. A.

- Biotropica* **30**, 275–285 (1998).
 11. Kinoshita, M., Sato, M. & Arikawa, K. *Naturwissenschaften* **84**, 199–201 (1997).
 12. Kelber, A. *Nature* **402**, 251 (1999).
Competing financial interests: declared none.

Social insects

Cuticular hydrocarbons inform task decisions

Social insect colonies are organized without central control, and must not only accomplish many tasks, such as foraging and nest construction, but must also respond to changing conditions by adjusting the number of workers performing each task^{1,2}. Here we use chemically treated, artificial ants to show that cuticular hydrocarbons, which differ according to task, are used by workers of the red harvester ant (*Pogonomyrmex barbatus*) to recognize the tasks of the ants that they encounter. Encounters with other ants thus inform a worker's decision on whether to perform a particular task.

A mature colony of the red harvester ant, a seed-eating desert species, consists of a single queen and 10,000–12,000 workers. We focused on two task groups: foragers, who collect food; and patrollers, who scout the foraging area each morning. If patrollers do not return safely, foragers will not leave the nest to search for seeds³. Nest-maintenance workers are active at the same time as patrollers and do not stimulate foraging⁴. A social-insect worker can become

active or switch task as conditions are altered — depending, for example, on the number of other workers who are currently engaged in a particular task^{5–7}.

Communication in social insects occurs mostly by chemical and tactile means⁸, with cuticular hydrocarbons often acting as recognition cues⁹. A harvester ant's task decisions depend on its interaction, by antennal contact, with ants at the nest entrance¹⁰ — ants in different task groups differ in their cuticular hydrocarbon profiles¹¹. Foragers, for example, spend more time outside the nest and so are exposed to warmer, drier conditions than nest-maintenance workers, who mostly stay inside. This causes the foragers to have higher ratios of *n*-alkanes to *n*-alkenes and branched alkanes in their cuticular hydrocarbon profiles¹².

For field experiments, we used nine mature colonies at a long-term study site near Rodeo, New Mexico, in the United States¹³. We first inhibited foraging by removing returning patrollers. After 30 min of inactivity, we mimicked the flow of returning patrollers by dropping glass beads (3 mm in diameter) that had been coated with one ant-equivalent of extract into the nest at a rate of one every 10 seconds. The coating on the beads consisted of patroller cuticular lipids, patroller hydrocarbons,

nest-maintenance hydrocarbons (which acted as a control for task specificity), or plain solvent (blank control). As a positive control for forager activity, we used live patrollers that were captured and then immediately returned to the nest. Cuticular lipids were extracted in 100% pentane for 10 min^{9,11} and hydrocarbons were purified from cuticular lipids by using column chromatography⁹.

The number of beads added to a nest was roughly equal to the number of patrollers collected. We then measured foraging activity by counting the number of active foragers outside the nest within 1 m of the entrance, every 10 min for 60 min. All colonies received each treatment in a random order; for each colony, we carried out one trial per day for five consecutive days. We normalized for variation among colonies in absolute forager number by dividing each mean number foraging per trial by the largest number of foragers ever observed for that colony.

Task-specific cuticular hydrocarbons from patrollers were sufficient to rescue foraging activity (Fig. 1). However, the behaviour is not a simple response to patroller extract alone. Our results, including preliminary data (not shown), indicate that in this patroller-mimic assay, all of the following are necessary to stimulate foraging activity: a one-ant equivalent concentration of hydrocarbon extract, location just inside the nest entrance, sequential presentation, and the time of day at which the colony is ready to begin foraging.

A brief encounter with a nestmate influences an ant's task decision because the encounter identifies the task of the other worker, cued by subtle features of other ants' hydrocarbon profiles. Encounters between ants thus provide information used for task allocation. These encounters in the aggregate produce a dynamic network that regulates the colony's behaviour.

Michael J. Greene, Deborah M. Gordon

Department of Biological Sciences, Stanford University, Stanford, California 94305, USA
 e-mail: greene@ants.stanford.edu

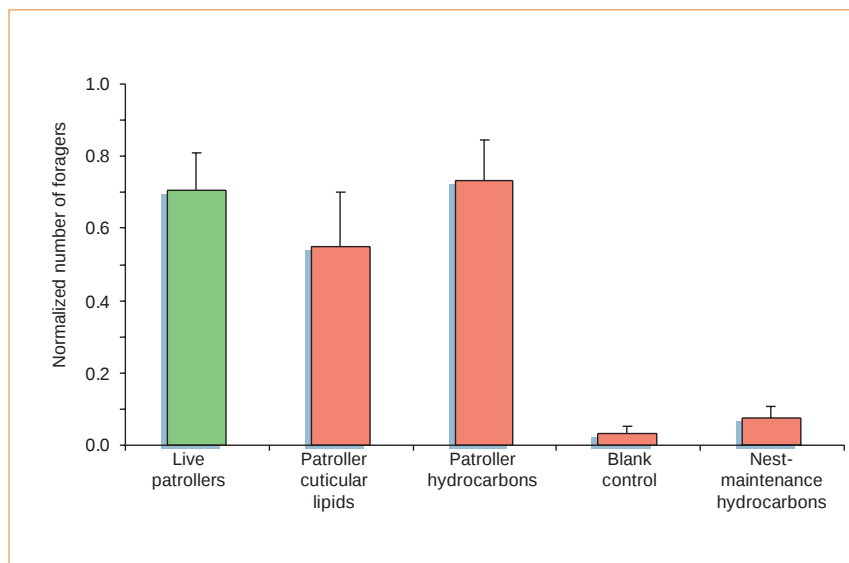


Figure 1 Task-specific cuticular hydrocarbons from patrollers are sufficient to rescue foraging activity in red harvester ants. The number of foraging ants (normalized; see text) leaving the nest is shown in response to live patrollers returning to the nest (green bars) or to different hydrocarbon-coated glass-bead ant mimics (red bars). Significantly more foragers emerged in response to live patrollers and to ant mimics treated with patroller cuticular lipids or patroller hydrocarbons than to mimics coated with blank control or nest-maintenance-worker hydrocarbons (repeated-measures analysis of variance: $F_{4,28} = 11.88$, $P < 0.0001$; $n = 9$). There was no significant difference in foraging-ant numbers among the returned live patrollers, patroller cuticular lipid and patroller hydrocarbon treatments, or between the blank and nest-maintenance hydrocarbon treatments (Tukey's post-hoc analysis). Data were transformed with an angular transformation (square-root of arcsine) for analysis.

- Gordon, D. M. *Nature* **380**, 121–124 (1996).
- Robinson, G. E. *Annu. Rev. Entomol.* **37**, 637–665 (1992).
- Gordon, D. M. *Am. Nat.* **159**, 509–518 (2002).
- Gordon, D. M. *Anim. Behav.* **35**, 833–843 (1987).
- Wilson, E. O. *Behav. Ecol. Sociobiol.* **16**, 89–98 (1984).
- Winston, M. L. & Fergusson, L. A. *Can. J. Zool.* **63**, 777–780 (1985).
- Gordon, D. M. *Anim. Behav.* **38**, 194–204 (1989).
- Hölldobler, B. & Wilson, E. O. *The Ants* (Belknap, Cambridge, Massachusetts, 1990).
- Nelson, D. R. & Blomquist, G. J. in *Biochemistry of Waxes* (eds Hamilton, J. & Christie, W. W.) Ch. 1 (Oily, Dundee, UK, 1995).
- Gordon, D. M. & Mehdiabadi, N. *Behav. Ecol. Sociobiol.* **45**, 370–377 (1999).
- Wagner, D. *et al.* *J. Chem. Ecol.* **24**, 2021–2037 (1998).
- Wagner, D., Tissot, M. & Gordon, D. M. *J. Chem. Ecol.* **27**, 1805–1819 (2001).
- Gordon, D. M. *Ants at Work: How an Insect Society is Organized* (Norton, New York, 2000).

Competing financial interests: declared none.

X-ray structure of a voltage-dependent K⁺ channel

Youxing Jiang*, Alice Lee, Jiayun Chen, Vanessa Ruta, Martine Cadene, Brian T. Chait & Roderick MacKinnon

Howard Hughes Medical Institute, Laboratory of Molecular Neurobiology and Biophysics, Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Voltage-dependent K⁺ channels are members of the family of voltage-dependent cation (K⁺, Na⁺ and Ca²⁺) channels that open and allow ion conduction in response to changes in cell membrane voltage. This form of gating underlies the generation of nerve and muscle action potentials, among other processes. Here we present the structure of KvAP, a voltage-dependent K⁺ channel from *Aeropyrum pernix*. We have determined a crystal structure of the full-length channel at a resolution of 3.2 Å, and of the isolated voltage-sensor domain at 1.9 Å, both in complex with monoclonal Fab fragments. The channel contains a central ion-conduction pore surrounded by voltage sensors, which form what we call 'voltage-sensor paddles'—hydrophobic, cationic, helix–turn–helix structures on the channel's outer perimeter. Flexible hinges suggest that the voltage-sensor paddles move in response to membrane voltage changes, carrying their positive charge across the membrane.

Many cells produce electrical impulses, known as action potentials, that propagate along their surface membrane. Action potentials travel quickly, and their arrival at a distant location initiates cellular processes such as the release of neurotransmitter molecules or the contraction of muscles¹. These electrical impulses are the means by which living cells transfer information over large distances in short time intervals, allowing information processing in the nervous system, movement and many other processes.

Fifty years ago, Hodgkin and Huxley presented the theory of the action potential, which included two key elements—that the membrane of a cell can undergo transient changes in its selective permeability to Na⁺ and K⁺ ions, and that the permeability changes depend on membrane voltage². These two elements create an interesting situation, because selective permeability to ions determines the membrane voltage, and the voltage determines the permeability. The feedback relationship between membrane voltage and ion permeability, combined with the 'electrical cable' properties of a cylindrical membrane, gave an accurate description of action potentials in squid giant axons.

These key elements of the Hodgkin–Huxley theory are embodied in a single family of protein molecules known as the voltage-dependent cation channels¹. This family includes K⁺, Na⁺ and Ca²⁺-selective members. When the pore of a voltage-dependent cation channel opens, it selectively conducts predominantly its namesake ion. The control of pore opening—the process known as gating—is dependent on the membrane voltage, hence the name 'voltage-dependent channel'.

How does a voltage-dependent channel 'sense' the membrane voltage and gate open as a function of its value? It is thought that charged amino acids called 'gating charges' move through the membrane electric field in association with pore opening, allowing membrane voltage to bias the equilibrium between closed and opened conformations^{3–5}. Gating charge movements can be measured as transient electrical currents associated with gating, separate from ion conduction³. In K⁺ channels, the gating charge per tetrameric channel corresponds to 12–14 electron charges (3.0–3.5 charges per subunit) crossing the entire membrane voltage difference^{6–8}. This large gating charge gives rise to a steep change in open probability as a function of membrane voltage.

All members of the voltage-dependent cation channel family

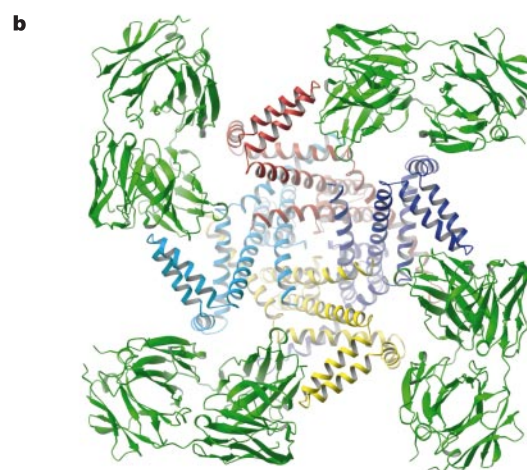
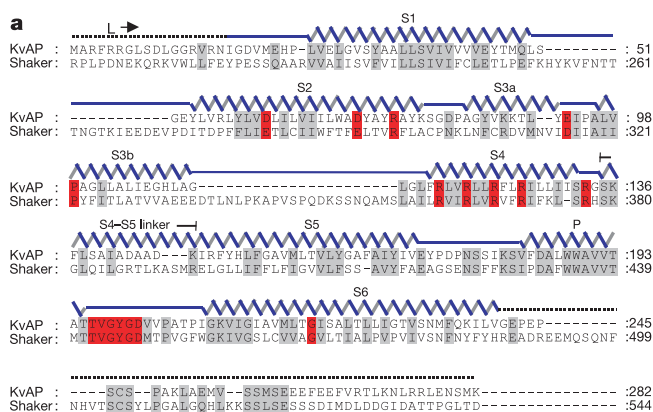


Figure 1 Structure of the KvAP channel. **a**, Sequence alignment of KvAP and the *Shaker* K⁺ channel. Regions of high homology are coloured grey, and residues known to be important for K⁺ selectivity and voltage-dependent gating are coloured red. Secondary structure elements show S1–S6 and P (pore) α -helices on the basis of the crystal structure. Disordered regions are shown as black dotted lines. The first five residues of KvAP were replaced by a single leucine (L) during expression. **b**, Structure of the KvAP channel (blue, yellow, cyan and red helical structures) bound to four Fab fragments (green), viewed down the four-fold axis from the intracellular side. This and all subsequent ribbon diagrams were generated with the program RIBBONS⁴².

* Present address: University of Texas Southwestern Medical Center, Department of Physiology, 5323 Harry Hines Blvd, Dallas, Texas 75390-9040, USA.

contain six hydrophobic segments per subunit, S1–S6 (refs 9–11). Four subunits (most often identical in K^+ channels, and linked together as homologous ‘domains’ in Na^+ and Ca^{2+} channels) surround a central ion-conduction pore. S5–S6 line the pore and determine ion selectivity, whereas S1–S4 form the voltage sensors. Certain charged amino acids within the voltage sensors, particularly the first four arginines in S4, account for most of the gating charge^{7,8}. Studies with thiol-reactive compounds^{12–15} and fluorescently labelled channels^{5,16,17} have documented conformational

changes in the region of S4 during gating, and form the basis of the prevailing model, in which S4 helices move along a path surrounded by protein, shielded from lipid^{18–20}.

We have studied the mechanism of voltage-dependent gating using biochemical, X-ray crystallographic and electrophysiological methods. Our studies focus on a voltage-dependent K^+ channel, KvAP, from the thermophilic archaebacteria *Aeropyrum pernix*. This channel is closely related in amino-acid sequence to, and exhibits all of the electrophysiological properties of, eukaryotic Kv

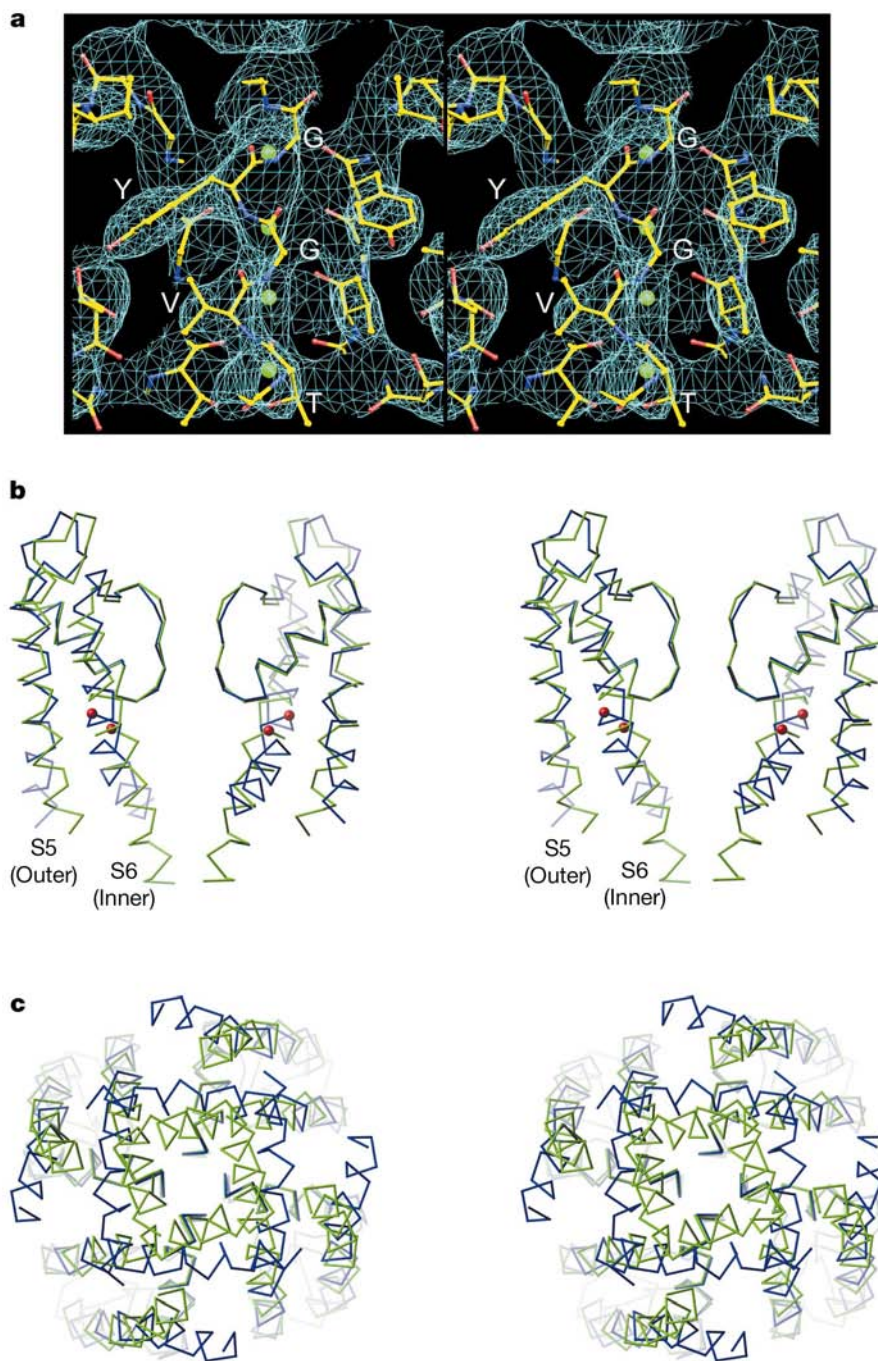


Figure 2 Stereo view of the KvAP pore and comparison with the KcsA K^+ channel. **a**, An electron density map at 1.0 σ contour (blue mesh) was calculated from the Fab model used in molecular replacement after rigid-body refinement. The refined model of the selectivity filter is shown with carbon, nitrogen and oxygen atoms coloured yellow, blue and red, respectively. The positions of four K^+ ions are shown as green spheres.

The α -carbon traces of the KvAP pore (blue) and the KcsA K^+ channel (green) are shown from the side in **b** and from the intracellular solution in **c**. Outer and inner helices are labelled S5 and S6, respectively. The glycine-gating hinges are shown as red spheres.

Table 1 Data and refinement statistics

Protein	Fab 6E1-KvAP	Fab 33H1-isolated voltage sensor
Space group	I422	C2
Data source*	CHESS A1	CHESS A1
Resolution (Å)	30–3.2	30–1.9
Completeness (%)	98.8 (95.2)	97.1 (92.8)
Redundancy†	5.5	2.9
R_{sym} (%)‡	8.0 (42.4)	5.8 (17.7)
I/σ	18 (2.2)	15.5 (5.8)
Unique reflections	22,476	56,778
Atoms refined	5,046 protein, 6 H ₂ O, 6 K, 7 Cd	4,365 protein, 403 H ₂ O
$R_{\text{work}}/R_{\text{free}}$ (%)§	25.6/29.9	23.1/25.1
r.m.s.d. of bond	1.44°/0.009 Å	1.48°/0.006 Å

*Data listed are those used for refinement. Numerous data sets were also collected at CHESS F1 and NSLS X25.

†Redundancy represents the ratio of the total number of measurements to the number of unique reflections.

‡ $R_{\text{sym}} = \sum |I_i - \langle I_i \rangle| / \sum I_i$, where $\langle I_i \rangle$ is the average intensity of symmetry-equivalent reflections.

§ R factor = $\sum |F(\text{obs}) - F(\text{cal})| / \sum F(\text{obs})$; 5% of the data that were excluded from refinement were used in the R_{free} calculation.

||r.m.s.d. of bond is the root-mean-square deviation of bond angle and length.

Numbers in parentheses are statistics for last resolution shell.

channels²¹. We describe the structure of KvAP and two conformations of its voltage sensor. The structure of the voltage sensor is very unexpected. In a companion paper²², we test mechanistic predictions of this structure to begin to understand how the channel senses membrane voltage to open its pore.

Structure determination

We examined many different bacterial K⁺ channels that contain amino-acid sequences corresponding to the voltage sensor. For six of these channels, we succeeded in expressing milligram quantities

of tetrameric channels, but in every case they aggregated into multiples of tetramers during concentration, and never crystallized. Even after reconstitution into lipid membranes at high protein-to-lipid ratios, rafts of closely packed channels did not form ordered arrays, which suggested that the voltage sensor might be a very mobile structure, making crystallization difficult. With this difficulty in mind, we initiated a search for voltage-dependent K⁺ channels in extreme thermophiles, hoping to identify an exceptionally stable family member.

Figure 1a shows the sequence of KvAP from *Aeropyrum pernix*²¹, together with the *Shaker* K⁺ channel sequence¹¹. Amino acids that are known to be crucial for ion selectivity and voltage-dependent gating are highlighted in red to emphasize the relatedness of these two voltage-dependent K⁺ channels. KvAP was expressed and purified, and remained as a mono-dispersed tetramer in a variety of detergents. During expression in *Escherichia coli*, we found it necessary to block K⁺ channels with 10 mM BaCl₂ to reduce toxicity to bacteria and to obtain properly folded protein²¹. For an unknown reason, the expressed KvAP channel contains a leucine in place of the first five amino acids encoded by the KvAP gene (see Methods).

Even the KvAP channel failed to crystallize under a wide range of conditions. To overcome the potential flexibility of the voltage-sensor regions, we reasoned that if we could 'hold' the channel with a Fab fragment attached to each of its four voltage sensors, we might favour crystallization. To this end, we raised monoclonal antibodies against the channel, and selected those that could bind to the voltage sensor (see Methods). These were produced in large quantity, and used (as Fab fragments) in crystallization trials. One of these (Fab 6E1) yielded crystals that diffracted X-rays to 3.2-Å Bragg spacings

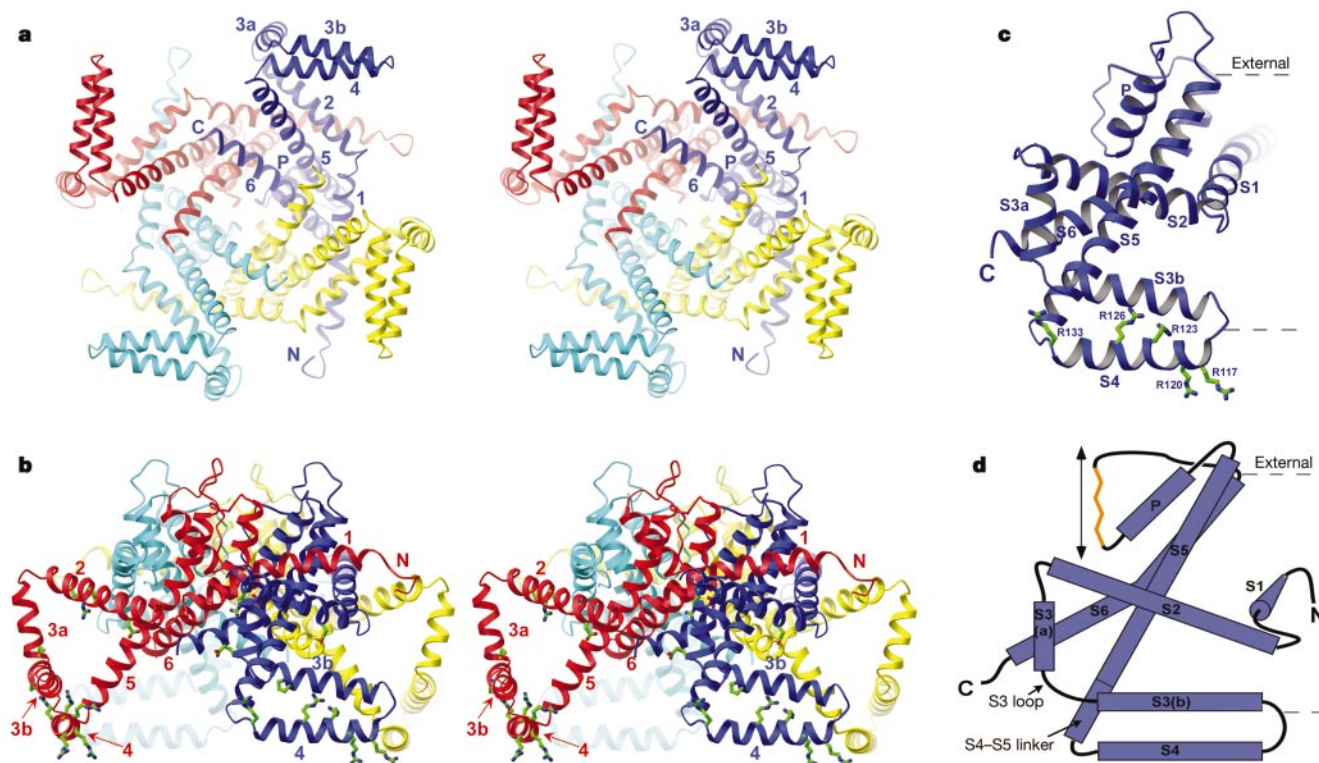


Figure 3 Architecture of the KvAP channel. **a**, Stereo image of the KvAP channel tetramer viewed from the intracellular side of the membrane. Each subunit is a different colour; helical elements of the blue subunit are labelled numerically for S1–S6 and with P for the pore helix. N and C mark the termini. **b**, Stereo image of the KvAP channel tetramer viewed from the side with the intracellular solution below. Views in **a** and **b** are related by a 90° rotation about a horizontal axis. Side chains of selected residues known to be involved

in voltage-dependent gating are shown on the blue and red subunits. Selected helical elements are labelled on the red and blue subunits. **c**, A single KvAP subunit viewed from the same perspective as in **b**. Conserved arginine residues on the voltage-sensor paddle are shown. **d**, A schematic diagram of the KvAP subunit topology is shown with an orange selectivity filter and an arrow to indicate the ion pathway. The demarcation between the S4–S5 linker and S5 is indicated by a black line.

at the synchrotron. Phases were determined by molecular replacement using a Fab search model, and after rigid-body refinement, they yielded a map with interpretable electron density in the channel region (Fig. 2a). Difference Fourier maps calculated with Hg derivative data from site-directed cysteine mutants were used to guide model building (see Methods). A complete model from amino acid 18 to 240 including all side chains was refined to a free residual of 29.9% (Table 1). The crystallographic solution reveals a four-fold symmetric tetramer with a Fab bound at the outer perimeter of each subunit (Fig. 1b).

The ion-conduction pore

Because of its high conservation among K^+ channels, the selectivity filter (sequence TVGYG) was the one region of the KvAP channel where we could predict what the structure ought to look like. As expected, the filter is essentially identical to that in KcsA^{23,24} (Fig. 2a). The filter features carbonyl oxygen atoms directed towards the pore to coordinate dehydrated K^+ ions, and hydrophobic side chains of valine and tyrosine directed into a hydrophobic core surrounding the filter to stabilize the main chain. Although certain amino-acid side chains within the hydrophobic core surrounding the filter are different from those found in KcsA, a comparison of α -carbon traces shows almost perfect superposition (Fig. 2b). The near invariance of the selectivity filter structure in the context of channels (KcsA and KvAP) with very different gating mechanisms is not surprising, because in both channels the filter serves the same function, to conduct K^+ ions rapidly and selectively.

The structures of the KvAP and KcsA pores deviate within the region of the intracellular membrane leaflet (Fig. 2b, lower half).

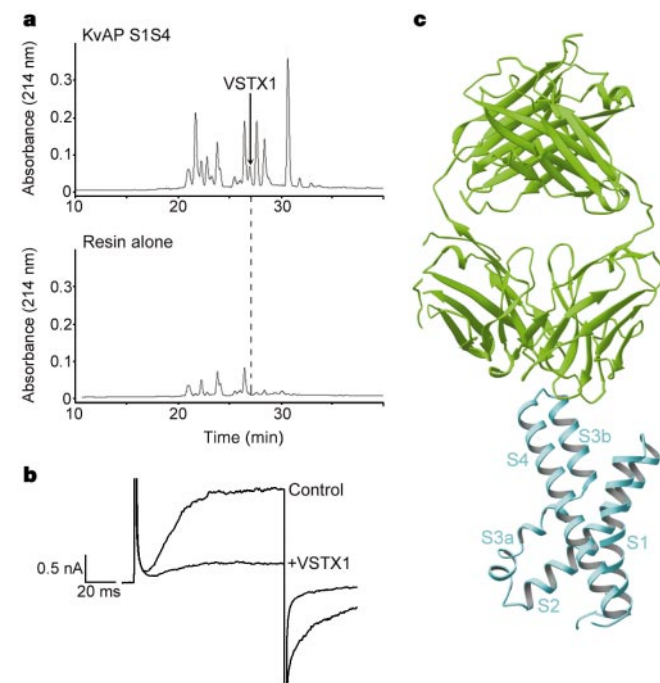


Figure 4 Functional and structural analysis of the isolated voltage-sensor domain. **a**, The isolated voltage sensor retains its ability to bind tarantula toxins that specifically inhibit voltage sensors. A voltage-sensor affinity column was constructed by saturating Co^{2+} affinity resin with the voltage-sensor protein. The column retained toxin components (top), including the known active VSTX1 (arrow), whereas a column formed from the Co^{2+} resin alone did not (bottom). Graphs show reverse-phase HPLC chromatograms of toxins eluted from the Co^{2+} columns. **b**, VSTX1 inhibits KvAP channel currents elicited by a +100 mV depolarization. **c**, Structure of the isolated voltage sensor (cyan) bound to Fab 33H1 (green).

The S6 helices (also known as inner helices²³), which line the ion pathway below the selectivity filter, begin to diverge most noticeably at a point referred to as the glycine-gating hinge (Fig. 2b, red spheres). The gating hinge was identified when comparing the structures of KcsA (a closed K^+ channel) and MthK (an opened K^+ channel), and observing that the inner helices seem to open like the aperture of a camera by bending at a glycine residue²⁵. This glycine is conserved at a specific location in the inner helix of many K^+ channels, including KvAP, and has been proposed to underlie gating²⁶. Bending of the inner helix at the glycine-gating hinge position of KvAP suggests that it may gate open by similar movements.

The KvAP channel is opened wider than the closed KcsA K^+ channel²³ (Fig. 2c)—nearly as wide as the opened MthK channel²⁵. We think that the KvAP channel has an opened pore because of crystal packing interactions, which pull the outer helices away from the pore axis; we will return to this point later when considering the mechanism by which the voltage sensor may exert force on the pore to open it. For now, we wish to highlight one important feature of the structural comparison in Fig. 2c: in both KvAP and KcsA, the adjacent inner (S6) and outer (S5) helices of each subunit maintain an approximately antiparallel relationship to one another. That is, the inner and outer helices seem to move together as a single unit. This observation raises the interesting possibility that the voltage sensor, which is attached directly to the outer helix, might open the pore by pulling the outer helices away from the central pore axis, causing the inner helices to follow.

The voltage sensors

The S1–S4 helices are attached to the pore as shown in Fig. 3a–c. A view down the four-fold axis from inside the cell shows that the S6 and S5 helices from each subunit encircle the ion pathway, as would be expected from the KcsA²³ and MthK²⁵ structures. Unexpectedly, however, S1 and S2 form the next concentric layer of helices outside S5, and S3 and S4 are located at the channel's outer perimeter (Fig. 3a). The voltage sensor has been described in the past as a set of four helical segments, on the basis of amino-acid sequence analysis. The structure shows that the traditional notation needs to be revised, as shown in Fig. 3. Contained within the traditional definition of the S3 helix there are two individual helices (S3a and S3b) separated by a loop (S3 loop). The second helix in S3, S3b, and the amino-terminal half of the traditionally defined S4 are packed tightly against each other, forming a helix–turn–helix structure that is mainly hydrophobic, with the notable exception of distributed arginine residues (Fig. 3b, c, green). We call this S3b–S4 unit the voltage-sensor paddle. The arginine residues on the paddle are highly conserved among voltage-dependent K^+ channels^{4,5}, and the first four have been shown to account for most of the gating charge movement that underlies opening and closing of the pore^{7,8}. The Fab 6E1 is bound at the tip of each paddle, to the turn connecting S3b to S4 (Fig. 1b).

Several aspects of this structure are unexpected, but a second structure (below) will help to explain why by showing that the voltage sensor is intrinsically flexible, and is capable of adopting more than one conformation. The most unexpected finding is the position of the S4 helices, which are near the intracellular membrane surface, perpendicular to the pore axis (Fig. 3). This finding apparently contradicts solid electrophysiological data showing that toxins and thiol-reactive compounds can react with the N-terminal portion of S4 from the extracellular side of the membrane, but never from the intracellular side^{13–15,27}. How can this position of S4 helices in the crystal structure be reconciled with the electrophysiological studies? The structure shows that the voltage-sensor paddles are flanked on their N-terminal limit by the S3 loop, and on their carboxy-terminal limit by a sharp turn at a glycine residue into S5, the first several turns of which correspond to the C-terminal portion of the traditionally defined S4, referred to here as the S4–S5 linker

(Fig. 3c, d). These connections should allow the voltage-sensor paddles to move rather freely with respect to the main body of the channel inside the membrane, and possibly account for the above-mentioned difficulty in crystallization. Could it be that the voltage-sensor paddles' loose attachment to the pore allowed the Fabs to extend the hinges and pull the paddles towards the cytoplasm to accommodate packing in the crystal? To address this question, we carried out further experiments.

We expressed a partial channel corresponding to S1–S4 (the isolated voltage sensor) in the absence of the ion-conduction pore. This is an unusual approach—to produce only part of an integral membrane protein—but the outcome is informative. The isolated voltage sensor is a monomer on gel filtration—not surprisingly, because the subunits of a K^+ channel tetramerize through the pore-forming S5, P (pore) and S6 helices, with S1–S4 located out on the perimeter (Fig. 3). We examined whether the isolated voltage sensor retains a native structure by asking whether it can recognize and bind to protein toxins that specifically inhibit voltage-sensor movements^{27,28}. By attaching the isolated voltage sensor to resin on a column, we found that it efficiently purifies toxins from tarantula venom (Fig. 4a), and that these toxins inhibit functional KvAP channels in an electrophysiological assay (Fig. 4b). Some of the toxins we have identified in this way are the same as those previously described, such as VSTX1 (ref. 21), whereas others are new members of the family of voltage-sensor toxins. For our immediate purpose, however, this assay indicates that the isolated voltage sensor retains native structure, otherwise it would not exhibit molecular specificity for protein toxins that have evolved to bind to the voltage sensor with high affinity.

We determined the structure of the isolated voltage sensor in a complex with a Fab at 1.9-Å resolution by X-ray crystallography (Table 1 and Fig. 4c). In this isolated domain structure, the voltage

sensor is not constrained by its attachment to the pore. We again solved the phase problem by molecular replacement using a Fab search model, and refined a complete atomic model to a free residual of 25.1%. The Fab (33H1) was raised against the full-length channel and selected on the basis of its binding to the voltage sensor (somewhere within S1–S4), as described in the Methods. We note that Fab 33H1 (in the isolated voltage-sensor structure) and Fab 6E1 (in the full-length channel structure) bind to precisely the same epitope, with the same relative orientation, at the tip of the voltage-sensor paddle (Figs 1b and 4c).

Figure 5a shows the structure of the isolated voltage sensor. It contains an arrangement of helices S1, S2, S3a, S3b and S4. An S3 loop between S3a and S3b, and a voltage-sensor paddle formed by S3b and the N-terminal half of S4, are immediately recognizable. Side-chain structures of highly conserved and functionally important voltage-sensor amino acids, highlighted red in the alignment in Fig. 1a, are well defined at 1.9-Å resolution and are shown in detail (Fig. 5a). The first four arginine residues (positions 117, 120, 123 and 126) on the voltage-sensor paddle are exposed to solvent in the crystal. These are the ones shown to carry most of the gating charge in the *Shaker* K^+ channel^{7,8}. The fifth S4 arginine (position 133) makes a salt bridge with aspartate 62, appearing to glue S4 and S2 together right beside the S3 loop. We note that arginine 133 in KvAP aligns most closely with arginine 377 in the *Shaker* K^+ channel S4 (Fig. 1a), and that *Shaker* K^+ channel function is destroyed when this residue is mutated⁵. A salt-bridge network involving arginine 76 in S2, aspartate 72 one helix turn away in S2, and glutamate 93 in S3a, seems to bridge S3a to the C-terminal half of S2. Lysine 136 in KvAP, which is conserved in the *Shaker* K^+ channel as lysine 380 (Fig. 1a) but not in all voltage-dependent K^+ channels, also participates in this salt-bridge network. The overall distribution of basic and acidic amino acids in the isolated voltage-sensor

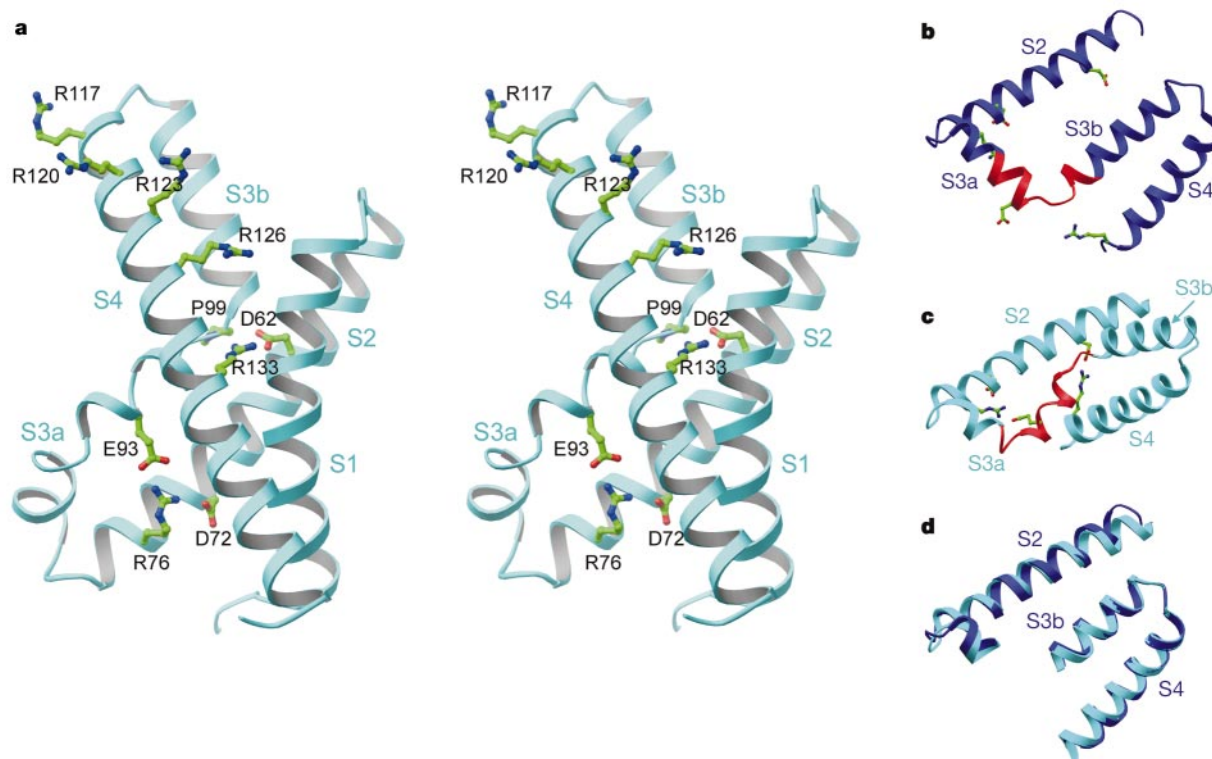
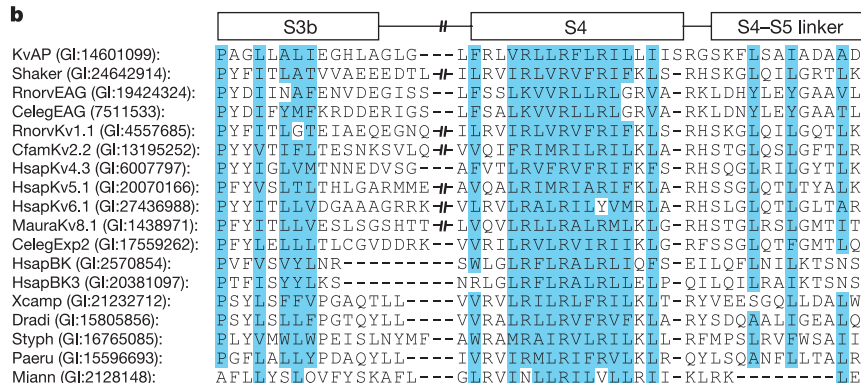


Figure 5 Structure of the isolated voltage sensor. **a**, Stereo view with helical elements labelled in cyan and selected amino acids labelled in black using the single-letter code. **b–d**, The voltage sensor from the beginning of S2 to the end of the voltage-sensor paddle is shown for the full-length KvAP channel (**b**), for the isolated voltage sensor (**c**),

and with S2 and the voltage-sensor paddle from both structures superimposed (**d**). Amino acids forming salt bridges in the isolated voltage sensor are shown, and the S3 loop is coloured red.

There are two major differences between the voltage sensor in the full-length structure and the isolated domain: first, S1 is folded back in the isolated domain (Fig. 5a; see below); and second, S4 is straight in the isolated domain (Fig. 5a) and bent in the full channel (Fig. 3), the bend occurring just after the voltage-sensor paddle. The reason for the S4 difference is clear: in the full-length channel, S4 bends at a glycine residue to redirect the helix into S5, whereas in the isolated voltage sensor, S4 is unrestrained and continues as a straight helix. It is significant that the amino-acid sequence corresponding to the N-terminal half of the traditionally defined S4, which forms the paddle and is structurally identical in both the full-length and

The isolated voltage sensor provides insight into why S4 is in an unexpected position, projecting beyond the cytoplasmic interface in the full-length channel crystal structure (Fig. 3). The full-length channel (Fig. 7a) is compared to the pore with the isolated voltage sensor (from the beginning of S2 to the end of the voltage-sensor paddle) docked onto it (Fig. 7b); docking was accomplished by placing S2 of the isolated voltage sensor into the location of S2 in the full-length channel. The comparison gives the immediate impression that the Fabs simply pulled the paddles down by extending their attachment through the S3 loop. The remaining gap between the voltage-sensor paddle and S5 could easily be accommodated by introducing a bend into the S4–S5 linker, which corresponds to three-and-a-half helical turns at the bottom of S5 (Fig. 7). This comparison leads us to conclude that the full-



the region corresponding to the paddle, and weaker conservation in the S4–S5 linker. Alignment was generated using ClustalW⁴³ followed by manual adjustment and exclusion of loops between S3b and S4 as indicated by slashes.

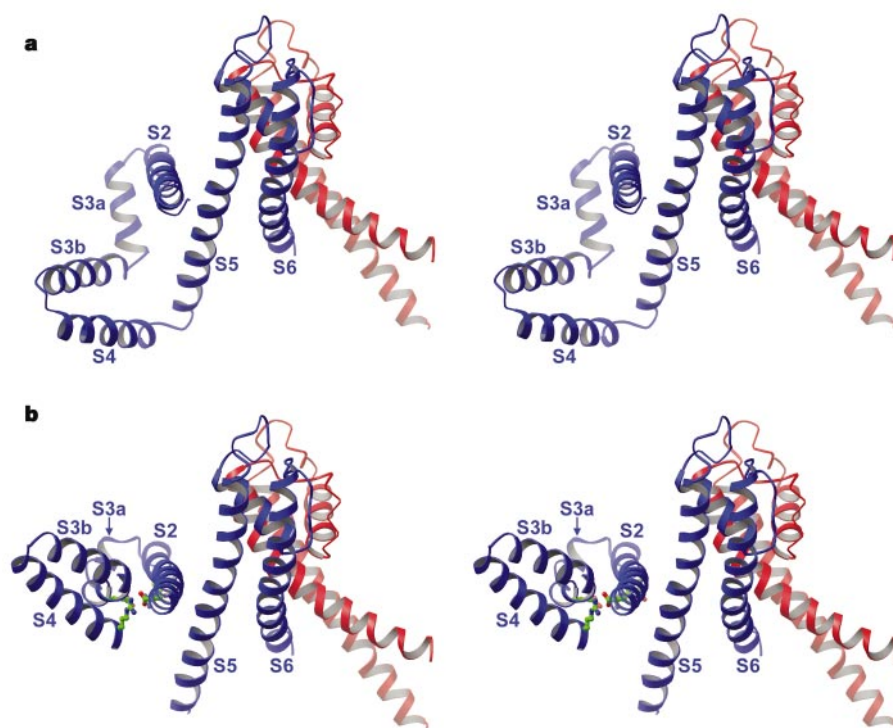


Figure 7 Effect of Fabs on voltage-sensor conformation. **a**, Stereo view of two subunits (blue and red) from the full-length channel viewed from the side with the intracellular solution below. **b**, Identical view of the pore (S5–S6) with the isolated voltage sensor

docked according to the position of S2 in the full-length channel. The Asp 62–Arg 133 salt bridge is shown.

length channel crystal structure is actually not very far from a membrane-bound conformation. In the accompanying paper, we provide further evidence for this conclusion by determining the positions of the paddles within the lipid membrane²².

Beyond the insight into how the Fabs have altered the voltage

sensors, comparison of two voltage-sensor structures offers information about their function. The voltage sensors would not so easily have conformed to crystal packing forces imposed by the Fabs if they did not contain a high degree of natural flexibility. Not only are the voltage-sensor paddles flexibly attached to S2 and S5, but S2 and S1 are loosely packed against the pore (Figs 3 and 7a). The entire voltage sensor would appear to float inside the membrane like a separate, mobile domain. Experiments by Lu and co-workers²⁹, who produced functional voltage-dependent K⁺ channels by splicing the *Shaker* K⁺ channel voltage sensor onto the KcsA pore (which does not interact naturally with a voltage sensor), support the same conclusion of loosely attached voltage-sensor domains; their chimaera channel would probably not function if the voltage sensors had to form a tight-fitting complementary interface with the pore.

In the *Shaker* K⁺ channel, glycosylation studies³⁰ and tethered blockers³¹ place the loop between S1 and S2 (which is larger than in KvAP by 19 amino acids; Fig. 1a) near the extracellular solution. The position of this loop in the middle of the membrane in KvAP could be a manifestation of the loose packing of S1 and S2 against the pore: we do not know to what extent the positions of these helices in the crystal deviate from that in the membrane, nor to what extent they move with gating. Furthermore, whether in the membrane S1 overlies an adjacent voltage sensor (as in the full-length channel; Fig. 3) or folds back (as in the isolated voltage sensor; Fig. 5) is unknown; a folded-back conformation would cause S1 to run straight into the cytoplasm. Despite these uncertainties, which are related to the voltage sensors' intrinsic flexibility, we know that S1 and S2 are against the pore and that the voltage-sensor paddles are at the protein–lipid interface.

The degree to which a protein's structure can be altered by crystal lattice forces depends on how rigid the protein is, and rigidity is related to function. This point can probably be best appreciated by

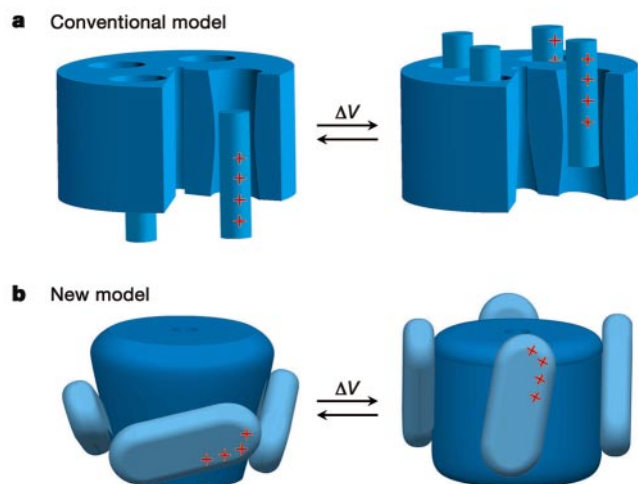


Figure 8 Hypothesis for gating charge movements. **a**, The conventional model of voltage-dependent gating posits that the gating charges are carried through the protein core by translations and/or rotations of S4 helices, which move independently of other protein segments within a 'gating pore' or 'canalculus'. **b**, In the new model, gating charges (red plus signs) could be carried through the membrane from inside (bottom) to outside (top) by movements of the voltage-sensor paddles against the lipid membrane, which in turn could open the pore.

examining the different regions of the KvAP channel itself. Near the selectivity filter, the pore is rigid, otherwise it would be difficult to explain the near-perfect superposition of KvAP and KcsA (Fig. 2b)—different K^+ channels in different crystals. The filter's rigidity makes sense in terms of its function, because it needs to be structurally constrained to discriminate between ions on the basis of small size differences. The voltage sensor (and cytoplasmic portions of the inner and outer pore-lining helices that gate the pore), by comparison, is flexible—capable of assuming different conformations (Fig. 5b–d). This property is related to its function as a voltage sensor: to sense voltage differences and open the pore, the voltage sensor has to move and rearrange itself within the membrane.

Summary

The KvAP voltage-dependent K^+ channel contains a canonical K^+ pore surrounded by voltage sensors. The voltage sensors feature S1 and S2 helices beside the pore, and voltage-sensor paddles on the outer perimeter. The voltage-sensor paddles are helix–turn–helix structures made from S3b and the N-terminal half of S4. Their amino-acid composition is mainly hydrophobic, with the important exception of S4 arginine residues. The paddles are movable with respect to the pore because of their flexible attachment via S3 loops. Movements of the paddles would be coupled to movements of S5 helices, and hence the pore, through the S4–S5 linker.

This structure has clear implications for how the voltage-sensor paddles might move in response to changes in the membrane voltage, and open the pore. In the crystal structure, the voltage-sensor paddles are located near the intracellular side of the channel, but electrophysiological studies using spider toxins and thiol-reactive compounds have shown that amino acids on the paddle can become exposed to the extracellular solution^{13–15,27}. These observations lead us to propose that the voltage-sensor paddles move across the membrane, carrying their cargo of gating charges through the electric field. Conventional models of voltage-dependent gating have posited that S4 helices move within their own protein-lined pathways, shielded from the lipid membrane by other parts of the protein (for example, S1, S2 and S3)^{18–20}, as depicted in Fig. 8a. Our results now lead us to a dramatically different picture, depicted in Fig. 8b. This new picture reminds us of Finkelstein's observations on voltage-dependent gating in colicin pore-forming toxins^{32,33}, in which positively charged peptides flip across the membrane. The concept that voltage-sensor paddles carry gating charges across the membrane at the protein–lipid interface suggests many new experiments. In an accompanying paper, we pursue some of these ideas by examining the positions of the voltage-sensor paddles in the membrane, and determining how far the paddles move when the pore opens in response to membrane depolarization²². □

Methods

Channel expression and purification

KvAP was expressed and purified as described²¹. DNA for the isolated voltage sensor encoding Met 1 to Lys 147 was cloned into a pQE60 expression vector (Qiagen) between *NcoI* and *BglII* sites with a thrombin cleavage site followed by a C-terminal hexahistidine sequence. Protein was expressed in *E. coli* XL1-Blue cells by induction (at $A_{600} \approx 1.0$) with 0.4 mM isopropyl- β -D-thiogalactopyranoside for 4 h at 37 °C. Cells were harvested and lysed in 50 mM Tris, pH 8.0, and 100 mM KCl, containing Leupeptin, Pepstatin, Aprotinin and phenylmethylsulphonyl fluoride (Sigma) to inhibit proteases. Protein was then extracted from the cell lysate for 3 h at room temperature in the above solution by adding 40 mM decylmaltoside (DM), purified on a Talon Co^{2+} affinity column (Clontech), and eluted with 50 mM Tris, pH 8.0, and 100 mM KCl and 300–400 mM imidazole. Protein was incubated overnight at room temperature (~ 21 °C) in the presence of 1 unit of thrombin per 2 mg of protein to cleave the hexahistidine tag, and then further purified on a Superdex-200 (10/30) column (Pharmacia) in a solution of 30 mM *n*-octyl- β -glucoside (β -OG), 20 mM Tris, pH 8.0, and 100 mM KCl. For both KvAP and the isolated voltage sensor, analysis by MALDI-TOF mass spectrometry (Voyager-STR; PerSeptive Biosystems) and N-terminal sequencing indicated that the N terminus undergoes modification during expression in which the first five residues of the encoded constructs are replaced with a single leucine residue. Site-directed mutagenesis using QuickChange

(Stratagene) was used to generate more than 30 cysteine mutants of KvAP for derivatization with mercuric compounds. In each of the mutants, single residues within the S1–S5 region were replaced by cysteine. Purification was the same as for the wild-type channel. Five sites (V43C, A71C, A84C, R126C, L162C) gave useful difference electron density to assist model building.

Toxin biochemistry

Isolated voltage sensor was expressed and purified as above, but without thrombin cleavage. Purified protein was applied to a 0.1 ml Co^{2+} column to saturate the resin. Venom from *Gramostola spatulata* (SpiderPharm) was diluted tenfold in 20 mM Tris, pH 8.0, 100 mM KCl and 10 mM DM, and applied (0.1 ml) to the column saturated with isolated voltage sensor and to a control column with resin alone. Both columns were washed extensively to minimize nonspecifically bound toxins, first in the above buffer, then in the above buffer with 15 mM imidazole. Remaining protein was eluted from both columns with 0.1 ml of 400 mM imidazole, and reduced with 50 mM dithiothreitol at 37 °C for 2 h to improve separation by reverse-phase high-performance liquid chromatography (HPLC). Equal volumes of eluted, reduced protein from the two columns were run on an Agilent 1100 Series HPLC with a C-18 reverse-phase 5- μ m, 80-Å column using a 2 min isocratic flow of 75% solution A (H_2O , 0.1% trifluoroacetic acid, TFA) and 25% solution B (90% acetonitrile, 10% H_2O , 0.1% TFA) followed by a 25–55% B gradient over 40 min. Peaks were collected and analysed by MALDI-TOF mass spectrometry. The peak corresponding to VSTX1 has the same retention time as purified, reduced VSTX1 run on the same gradient.

Fab complex preparation

Monoclonal antibodies (mouse immunoglobulin) against KvAP were raised using standard procedures²⁴. ELISA and western blot analysis were used to identify positive clones. Antibodies that recognized both KvAP and the isolated voltage sensor were selected. IgGs from mouse hybridoma cell culture supernatant were purified using a Protein A column (Bio-Rad). Fab fragments were generated by papain proteolysis (Worthington) and purified by Q-Sepharose chromatography. The sequences of the DNA encoding the variable region of Fabs were obtained as described²⁴. KvAP (in DM) and the isolated voltage sensor (in β -OG) were mixed with Fabs (6E1 and 33H1, respectively) and concentrated. The complexes were purified on a Superdex-200 (10/30) column in a solution of 30 mM β -OG, 20 mM Tris, pH 8.0, and 100 mM KCl. The purified complexes were concentrated to about 10 mg ml^{−1} using a 50-kDa molecular-mass cutoff Centricon (Amicon) for crystallization.

Crystallization

Crystals were grown by sitting-drop vapour diffusion at 20 °C by mixing equal volumes of protein and reservoir solutions. Fab 6E1–KvAP complex crystals were grown over a reservoir solution of 16–20% polyethylene glycol 400 (PEG400) or polyethylene glycol 350 monoethyl ether (PEGMME350), 150–200 mM CdCl₂ and 100 mM sodium acetate, pH 5.0. Crystals were of space group *I*422, with cell dimensions $a = b = 189.4$ Å, $c = 150.5$ Å, $\alpha = \beta = \gamma = 90^\circ$, and contained one KvAP subunit and one Fab in the asymmetric unit. Cryoprotection was achieved by raising the concentration of PEG400 or PEGMME350 in the reservoir to about 40%, after which crystals were frozen in liquid propane. Fab 33H1-isolated voltage-sensor complex crystals were grown over a reservoir solution of 20% PEG4000, 50 mM HEPES, pH 7.5. Crystals were of space group *C*2, with cell dimensions $a = 264.7$ Å, $b = 61.6$ Å, $c = 46.1$ Å, $\alpha = \gamma = 90^\circ$, $\beta = 90.35^\circ$, and contained one voltage sensor and one Fab in an asymmetric unit. Crystals were gently moved through paratone-N oil to remove excess solution surrounding the crystal, and were flash frozen under a cold nitrogen stream (-180 °C).

Structure determination

All data were collected at -180 °C under a nitrogen stream. For the Fab 6E1–KvAP complex crystals, multiple data sets (including those of cysteine mutants soaked in mercuric solution) were collected at Cornell High Energy Synchrotron Source (CHESS) A1 and F1, and National Synchrotron Light Source (NSLS) X25 beam lines. A crystal with a KvAP mutant (Tyr 46 being replaced by Cys) gave the best diffraction (3.2 Å). Data from Fab 33H1-isolated voltage-sensor domain crystals were collected at the CHESS A1 beam line. All data were indexed and integrated using DENZO and SCALEPACK³⁵, and processed with CCP4 programs³⁶. The Fab 6E1–KvAP structure was determined by molecular replacement using the program AMoRe³⁷ in CCP4 using a published Fab structure³⁸ (Protein Data Bank code 1BAF) as the search model. Because of the flexible connection between the Fab variable region (Fv) and constant region (Fc), these two units were used separately in the rotation and translation searches. After rigid-body refinement of the Fab fragment, a calculated electron density map using the phases from the Fab gave continuous electron density in the KvAP channel region. The cysteine mutants failed to provide useful phasing information owing to non-isomorphism; however, five sites (above) provided difference electron density maps to assist model building. The structure of the Fab 33H1-isolated voltage sensor was also determined by molecular replacement as described above. The complete models for both structures were built using O³⁹ and refined with CNS (Crystallography & NMR System)⁴⁰ by iterative cycles of simulated annealing and model rebuilding. The Fab 6E1–KvAP structure was refined against data to 3.2 Å to crystallographic and free residuals of 25.6% and 29.9% (using 5% of reflections), respectively. The final KvAP model contained residues 18–240. The N-terminal 17 residues and C-terminal 42 residues were disordered. Based on the presence of CdCl₂ and KCl in the crystallization solutions, seven Cd²⁺ and six K⁺ were modelled in the refined structure. The Fab 33H1-isolated voltage-sensor structure was refined against data to 1.9 Å with crystallographic and free residuals of 23.1% and 25.1% (using 5% of reflections), respectively. The first 19 residues of the isolated voltage sensor were disordered.

Electrophysiology

KvAP channels were reconstituted from DM into lipid vesicles as described⁴¹, to give a final protein concentration of 200 µg ml⁻¹. Planar lipid bilayers of 1-palmitoyl-2-oleoyl phosphatidylethanolamine (POPE; 15 mg ml⁻¹) and 1-palmitoyl-2-oleoyl phosphatidylglycerol (POPG; 5 mg ml⁻¹) in decane were painted over a 300 µm hole in a polystyrene partition separating the internal and external solutions. To induce fusion of channel-containing vesicles, solution on the side to which vesicles were added (*cis*) contained 150 mM KCl, 10 mM HEPES, pH 7.0, and the opposite side (*trans*) contained 15 mM KCl, 10 mM HEPES, pH 7.0. After the appearance of channels in the membrane, ion concentration on the *trans* side was increased to 150 mM KCl. VSTX1 was purified as described²¹.

Received 19 February; accepted 11 March 2003; doi:10.1038/nature01580.

1. Hille, B. *Ion Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 2001).
2. Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**, 500–544 (1952).
3. Armstrong, C. M. & Bezanilla, F. Charge movement associated with the opening and closing of the activation gates of the Na⁺ channels. *J. Gen. Physiol.* **63**, 533–552 (1974).
4. Sigworth, F. J. Voltage gating of ion channels. *Q. Rev. Biophys.* **27**, 1–40 (1994).
5. Bezanilla, F. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* **80**, 555–592 (2000).
6. Schoppa, N. E., McCormack, K., Tanouye, M. A. & Sigworth, F. J. The size of gating charge in wild-type and mutant Shaker potassium channels. *Science* **255**, 1712–1715 (1992).
7. Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K⁺ channel. *Neuron* **16**, 1159–1167 (1996).
8. Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K⁺ channel. *Neuron* **16**, 1169–1177 (1996).
9. Noda, M. *et al.* Primary structure of *Electrophorus electricus* sodium channel deduced from cDNA sequence. *Nature* **312**, 121–127 (1984).
10. Tanabe, T. *et al.* Primary structure of the receptor for calcium channel blockers from skeletal muscle. *Nature* **328**, 313–318 (1987).
11. Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, L. Y. & Jan, Y. N. Sequence of a probable potassium channel component encoded at *Drosophila*. *Science* **237**, 770–775 (1987).
12. Yang, N. & Horn, R. Evidence for voltage-dependent S4 movement in sodium channels. *Neuron* **15**, 213–218 (1995).
13. Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the Shaker K⁺ channel S4. *Neuron* **16**, 387–397 (1996).
14. Yusaf, S. P., Wray, D. & Sivaprasadarao, A. Measurement of the movement of the S4 segment during the activation of a voltage-gated potassium channel. *Pflugers Arch.* **433**, 91–97 (1996).
15. Baker, O. S., Larsson, H. P., Mannuzzu, L. M. & Isacoff, E. Y. Three transmembrane conformations and sequence-dependent displacement of the S4 domain in Shaker K⁺ channel gating. *Neuron* **20**, 1283–1294 (1998).
16. Mannuzzu, L. M., Moronne, M. M. & Isacoff, E. Y. Direct physical measure of conformational rearrangement underlying potassium channel gating. *Science* **271**, 213–216 (1996).
17. Cha, A. & Bezanilla, F. Characterizing voltage-dependent conformational changes in the Shaker K⁺ channel with fluorescence. *Neuron* **19**, 1127–1140 (1997).
18. Horn, R. Coupled movements in voltage-gated ion channels. *J. Gen. Physiol.* **120**, 449–453 (2002).
19. Gandhi, C. S. & Isacoff, E. Y. Molecular models of voltage sensing. *J. Gen. Physiol.* **120**, 455–463 (2002).
20. Bezanilla, F. Voltage sensor movements. *J. Gen. Physiol.* **120**, 465–473 (2002).
21. Ruta, V., Jiang, Y., Lee, A., Chen, J. & MacKinnon, R. Functional analysis of an archaeobacterial voltage-dependent K⁺ channel. *Nature* **422**, 180–185; advance online publication, 2 March 2003 (doi:10.1038/nature01473).
22. Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. The principle of gating charge movement in a voltage-dependent K⁺ channel. *Nature* **423**, 42–48 (2003).
23. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
24. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and

hydration revealed by a K⁺ channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).

25. Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
26. Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
27. Swartz, K. J. & MacKinnon, R. Mapping the receptor site for hanatoxin, a gating modifier of voltage-dependent K⁺ channels. *Neuron* **18**, 675–682 (1997).
28. Swartz, K. J. & MacKinnon, R. Hanatoxin modifies the gating of a voltage-dependent K⁺ channel through multiple binding sites. *Neuron* **18**, 665–673 (1997).
29. Lu, Z., Klem, A. M. & Ramu, Y. Ion conduction pore is conserved among potassium channels. *Nature* **413**, 809–813 (2001).
30. Santacruz-Tolosa, L., Huang, Y., John, S. A. & Papazian, D. M. Glycosylation of Shaker potassium channel protein in insect cell culture and in *Xenopus* oocytes. *Biochemistry* **33**, 5607–5613 (1994).
31. Blaustein, R. O., Cole, P. A., Williams, C. & Miller, C. Tethered blockers as molecular ‘tape measures’ for a voltage-gated K⁺ channel. *Nature Struct. Biol.* **7**, 309–311 (2000).
32. Slatin, S. L., Qiu, X. Q., Jakes, K. S. & Finkelstein, A. Identification of a translocated protein segment in a voltage-dependent channel. *Nature* **371**, 158–161 (1994).
33. Qiu, X. Q., Jakes, K. S., Kienker, P. K., Finkelstein, A. & Slatin, S. L. Major transmembrane movement associated with colicin Ia channel gating. *J. Gen. Physiol.* **107**, 313–328 (1996).
34. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1989).
35. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
36. Collaborative Computational Project, N.4 The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
37. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
38. Brunger, A. T., Leahy, D. J., Hynes, T. R. & Fox, R. O. 2.9 Å resolution structure of an anti-dinitrophenyl-spin-label monoclonal antibody Fab fragment with bound hapten. *J. Mol. Biol.* **221**, 239–256 (1991).
39. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
40. Brunger, A. T. *et al.* Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
41. Heginbotham, L., LeMasurier, M., Kolmakova-Partensky, L. & Miller, C. Single *Streptomyces lividans* K⁺ channels. Functional asymmetries and sidedness of proton activation. *J. Gen. Physiol.* **114**, 551–560 (1999).
42. Carson, M. Ribbons. *Methods Enzymol.* **277**, 493–505 (1997).
43. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

Acknowledgements We thank members of the MacKinnon laboratory for assistance at many stages of this project over five years, M. Long for biochemistry at early stages, J. Lee for studies of Ba²⁺ on cell growth, D. Wang for teaching us to use the electron microscope, F. Weis-Garcia for assistance with monoclonal antibodies, the staff at CHESS A1 and F1 and NSLSX25, O. Andersen and D. Gadsby for manuscript critique, and R. Mohan for graphic work. This work was supported in part by a National Institutes of Health (NIH) grant to R.M., and by the National Center for Research Resources, NIH to B.T.C. V.R. is supported by a National Science Foundation Graduate Student Research Fellowship, and R.M. is an Investigator in the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M. (mackinn@rockvax.rockefeller.edu). Coordinates have been deposited in the Protein Data Bank under accession codes 1ORQ and 1ORS.

The principle of gating charge movement in a voltage-dependent K⁺ channel

Youxing Jiang*, Vanessa Ruta, Jiayun Chen, Alice Lee & Roderick MacKinnon

Howard Hughes Medical Institute, Laboratory of Molecular Neurobiology and Biophysics, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

The steep dependence of channel opening on membrane voltage allows voltage-dependent K⁺ channels to turn on almost like a switch. Opening is driven by the movement of gating charges that originate from arginine residues on helical S4 segments of the protein. Each S4 segment forms half of a 'voltage-sensor paddle' on the channel's outer perimeter. Here we show that the voltage-sensor paddles are positioned inside the membrane, near the intracellular surface, when the channel is closed, and that the paddles move a large distance across the membrane from inside to outside when the channel opens. KvAP channels were reconstituted into planar lipid membranes and studied using monoclonal Fab fragments, a voltage-sensor toxin, and avidin binding to tethered biotin. Our findings lead us to conclude that the voltage-sensor paddles operate somewhat like hydrophobic cations attached to levers, enabling the membrane electric field to open and close the pore.

Voltage-dependent K⁺ channel opening follows a very steep function of membrane voltage¹. To allow channels to switch to the open state, gating charges—charged amino acids on the channel protein—move within the membrane electric field to open the pore^{1–3}. The crystal structure of KvAP, a voltage-dependent K⁺ channel, suggests how these gating charge movements might occur⁴. Four arginine residues are located on a predominantly hydrophobic helix–turn–helix structure called the voltage-sensor paddle. One paddle on each subunit is present at the outer perimeter of the channel. By moving across the membrane near the protein–lipid interface, the paddles could carry the arginine residues through the electric field, coupling pore opening to membrane voltage. We test this hypothesis for the movement of gating charges by estimating the positions of the voltage-sensor paddles inside the membrane when the channel is closed and opened.

Using Fabs and a toxin to detect paddle motions

Two monoclonal Fabs, 6E1 and 33H1, were used to crystallize and determine the structures of the full-length KvAP channel and the isolated voltage sensor, respectively⁴. Both Fabs were found attached to the same epitope on the tip of the voltage-sensor paddle between S3b and S4. We used these same Fabs to examine the position of the voltage-sensor paddles when the KvAP channel functions in lipid membranes (Fig. 1a), and to assess whether they change their position when the channel gates open in response to membrane depolarization. Figure 1b shows that both the 6E1 and 33H1 Fabs inhibit channel function when applied to the external solution. By contrast, neither Fab affected channel function from the internal solution.

Inhibition by external Fabs requires membrane depolarization (Fig. 1c). When the 33H1 Fab is added to the external chamber while the channel is held closed at –100 mV for 10 min (Fig. 1c, interval between the black and red data points), no inhibition is observed. The slightly larger current of the first red data point (20-min point) reflects recovery from a small amount of steady-state inactivation of channels occurring during the control pulse period (0–10-min data points)⁵. The important point, however, is that inhibition of current is detectable only on the second pulse following the addition of the Fab, as if the channel has first to open in order for the Fabs to bind. Inhibition progresses as the membrane is repeatedly depolarized. We also observe gradual recovery from

inhibition if, once the Fabs are bound, the membrane is held at a negative voltage for a prolonged period (30–40-min interval), implying that negative membrane voltages destabilize the interactions between channels and Fabs, causing the Fabs to dissociate. The same properties of inhibition are observed for Fab 6E1. Based on these results, we conclude that the voltage-sensor paddles must remain inaccessible as long as the channel is held closed by the negative membrane voltage, and that the entire epitope (two helical turns of S3b and one turn of S4) becomes exposed to the external side in response to depolarization.

Why do the Fabs inhibit the channel? If they bind to the voltage-sensor paddles from the external side when the membrane is depolarized, why do the Fabs not simply hold the channel permanently open? Inhibition can be explained by the fact that the KvAP channel, like most voltage-dependent K⁺ channels, inactivates⁵. That is, its pore stops conducting ions spontaneously during prolonged depolarizations of the membrane, even though the voltage sensors remain in their open conformation. Inactivation could also occur when the Fabs bind and hold the voltage sensors in their open conformation, thus explaining inhibition.

It is interesting that tarantula venom toxins also bind to residues on S3 and S4 (ref. 6), which we now know to be on the voltage-sensor paddles⁴. On addition of an approximately half-inhibitory concentration of the tarantula venom toxin VSTX1 to the external side of KvAP, we observe that the rate of inhibition is faster if the membrane is depolarized at a higher frequency (Fig. 1d). Therefore, the toxins, like the Fabs, require membrane depolarization.

We conclude from experiments with Fab fragments and a tarantula venom toxin that the voltage-sensor paddles are exposed to the extracellular solution during membrane depolarization (positive inside) but not during hyperpolarization (negative inside). In view of the KvAP crystal structure⁴, these results imply that the voltage-sensor paddles can move a large distance across the lipid membrane. How deep inside the membrane do the paddles sit when the membrane is hyperpolarized, and how far do they move when the channel opens?

Using biotin and avidin to measure paddle motions

Finkelstein and co-workers used avidin binding to biotinylated colicin to examine the movements of its components across the lipid bilayer^{7–9}. We subjected the KvAP channel to a similar analysis. The idea behind the experiments is outlined in Fig. 2a. We introduce cysteine residues at specific locations in the channel, biotinylate the cysteine, reconstitute channels into planar lipid membranes, and

* Present address: University of Texas Southwestern Medical Center, Department of Physiology, 5323 Harry Hines Blvd, Dallas, Texas 75390-9040, USA.

then determine whether avidin binds from the internal or external side, and whether binding depends on membrane depolarization. Wild-type KvAP channels contain a single cysteine on the carboxy-terminus, which was mutated to serine (without affecting function) to work with a channel without background cysteine residues.

The crystal structure of avidin and the chemical structure of biotin attached to its linker are shown in Fig. 2b. Biotin binds within a deep cleft inside the core of avidin, a rigid protein molecule¹⁰. The atom on the biotin molecule to which the linker is attached is 7 Å beneath the surface of avidin (Fig. 2b). Therefore, when avidin binds to a biotinylated cysteine on the channel, the distance from the cysteine α -carbon to the surface of avidin is 10 Å; this is the pertinent linker length from the α -carbon to avidin. Since avidin is large (the tetramer is a 57-kDa protein), it cannot fit into clefts on the channel and therefore cannot penetrate the surface. The important point is that, for avidin to bind to a tethered biotin, the cysteine α -carbon has to be within 10 Å of the bulk aqueous solution on either side of the membrane. Applying this restraint, we used linked biotin and avidin as a molecular ruler to measure positions of the voltage-sensor paddles.

Detection of avidin binding to biotinylated channels depends on the demonstration of a functional effect when avidin 'grabs' tethered biotin. Control experiments and examples are illustrated in Fig. 2c. Our convention for representing data will be black, red and blue traces corresponding to control (no avidin), external and internal avidin, respectively. Biotinylated wild-type channels are not affected by external avidin and show a small reduction in current when avidin is applied to the internal solution. On the basis of

protein gel assays, we conclude that the wild-type channels do not contain detectable biotin on them after the biotinylation procedure (not shown). Three examples of biotinylated cysteine mutant channels are shown. The G112C mutant is inhibited completely by external but not internal avidin; the small reduction by internal avidin is similar to the wild-type control. The I127C mutant is inhibited completely by internal but not by external avidin. The L103C mutant shows an outcome different from complete inhibition: external avidin reduces the current but does not abolish it, and changes the kinetics of current activation. Another outcome that is observed at certain positions is partial inhibition by avidin without changing the gating kinetics (Fig. 3). These cases probably reflect incomplete biotinylation of buried cysteine residues. We excluded the possibility that 'no effect' (that is, internal avidin on G112C or external avidin on I127C) represents 'silent' avidin binding by adding avidin first to the 'no effect' side of the membrane and then to the opposite side (not shown). Because avidin binding to biotin is essentially irreversible, silent binding to one side should protect from binding to the other, but this was never observed. Thus, 'no effect' signifies no binding.

Data from a scan of the voltage-sensor paddles from amino-acid position 101 on S3b to position 127 on S4 are shown in Fig. 3. Positions were mutated individually to cysteine, biotinylated and studied in planar lipid membranes. Certain biotinylated mutants showed altered gating even before avidin addition (for example, valine 119), or appeared to be less abundant in the membrane (for example, leucine 121). But in all cases, channels could be held closed at negative membrane voltages (typically -100 mV) and opened by

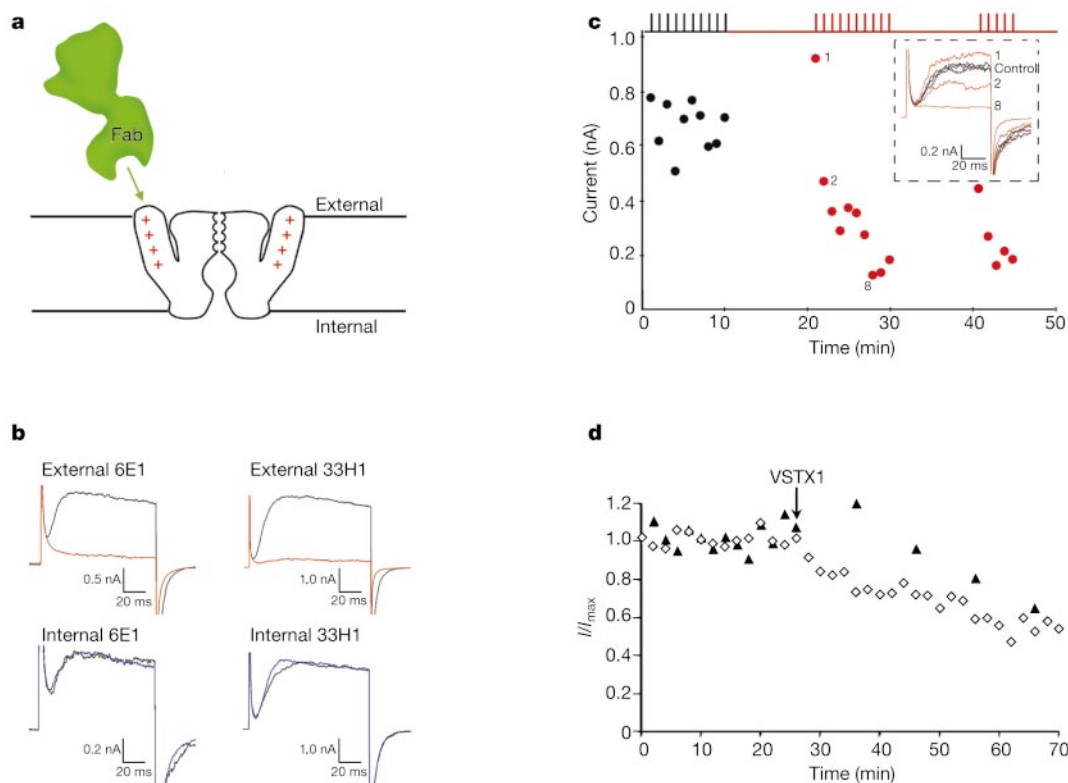


Figure 1 Inhibition of KvAP channels by Fabs and a tarantula venom toxin that bind to the voltage-sensor paddles. **a**, Experimental strategy: Fab (green) is added to the external or internal side of a planar lipid membrane to determine whether the epitope is exposed. **b**, Fabs used in crystallization (6E1 and 33H1) inhibit from the external (red traces) but not the internal (blue traces) side of the membrane. Currents before (black traces) and after the addition of about 500 nM Fab (red and blue traces) were elicited by membrane depolarization to 100 mV from a holding voltage of -100 mV. **c**, Fab 33H1 binds to the voltage-sensor paddle only when the membrane is depolarized. Current elicited by

depolarization from -100 mV to 100 mV at times indicated by the stimulus trace (above) in the absence (black symbols) or presence of 500 nM Fab (red symbols) is presented as a function of time. Selected current traces corresponding to the numbered symbols are shown (inset). **d**, VSTX1 binds from the external side only when the membrane is depolarized. Currents, normalized to the average control value, were elicited by a 100-ms depolarization to 100 mV every 120 s (diamonds) or every 600 s (triangles). VSTX1 (30 nM) was added to the external solution at the point indicated by the arrow.

membrane depolarizations (typically 100 mV for 200 ms) every 120 s. To compensate for mutations that shifted the voltage-activation curve, we sometimes held the membrane as negative as -140 mV, or used opening depolarizations as positive as 200 mV. Avidin bound to the tethered biotin and affected channel function at all positions studied. In many cases, avidin caused complete inhibition, and in some cases partial inhibition with altered kinetics. Partial inhibition at certain sites (for example, leucine 110 and leucine 122) can be explained on the basis of incomplete biotinylation because the side chain is buried within the 'core' of the voltage-sensor paddle between S3b and S4. This finding is consistent with the proposal that S3b and S4 move together as a voltage-sensor paddle unit.

We are most interested in whether a particular position on the voltage-sensor paddle allows binding to avidin from the external or internal solution. Amino acids on the paddle are colour-coded according to the membrane side from which avidin bound: red, outside; blue, inside; yellow, both. For avidin binding from the external side, we examined whether membrane depolarization is required, by studying the effect of depolarization frequency on the rate of channel inhibition by avidin (Fig. 4). Inhibition from the external solution required depolarization: higher frequencies gave

higher rates of inhibition. In other words, the voltage-sensor paddles can be protected from external avidin binding by keeping the membrane at negative voltages, and the paddles are exposed to external avidin at positive voltages. This was the case for all external positions (Fig. 3, red and yellow).

However, protection from inhibition by external avidin, Fabs and a voltage-sensor toxin, by holding the membrane at negative voltages, is incomplete. In particular, if the wait at negative voltages is long enough, inhibition occurs but at a low rate (Fig. 4). This finding is explained on the basis of thermal fluctuations of the voltage sensors. Probably four voltage-sensor paddles have to move to open the pore^{1,11–13}. But individual paddle movements must occasionally occur even at negative voltages. These sensor movements underlying incomplete protection are consistent with gating currents preceding pore opening², and longer delays before pore opening when the membrane is depolarized from more negative holding voltages, a phenomenon known as the Cole–Moore effect¹⁴.

Biotin molecules tethered at two positions were captured by avidin from both sides of the membrane (121 and 122; Fig. 3, yellow). At these positions, inhibition was complete from either side alone, because subsequent addition of avidin to the opposite side caused no further inhibition. Thus, the dual accessibility cannot be

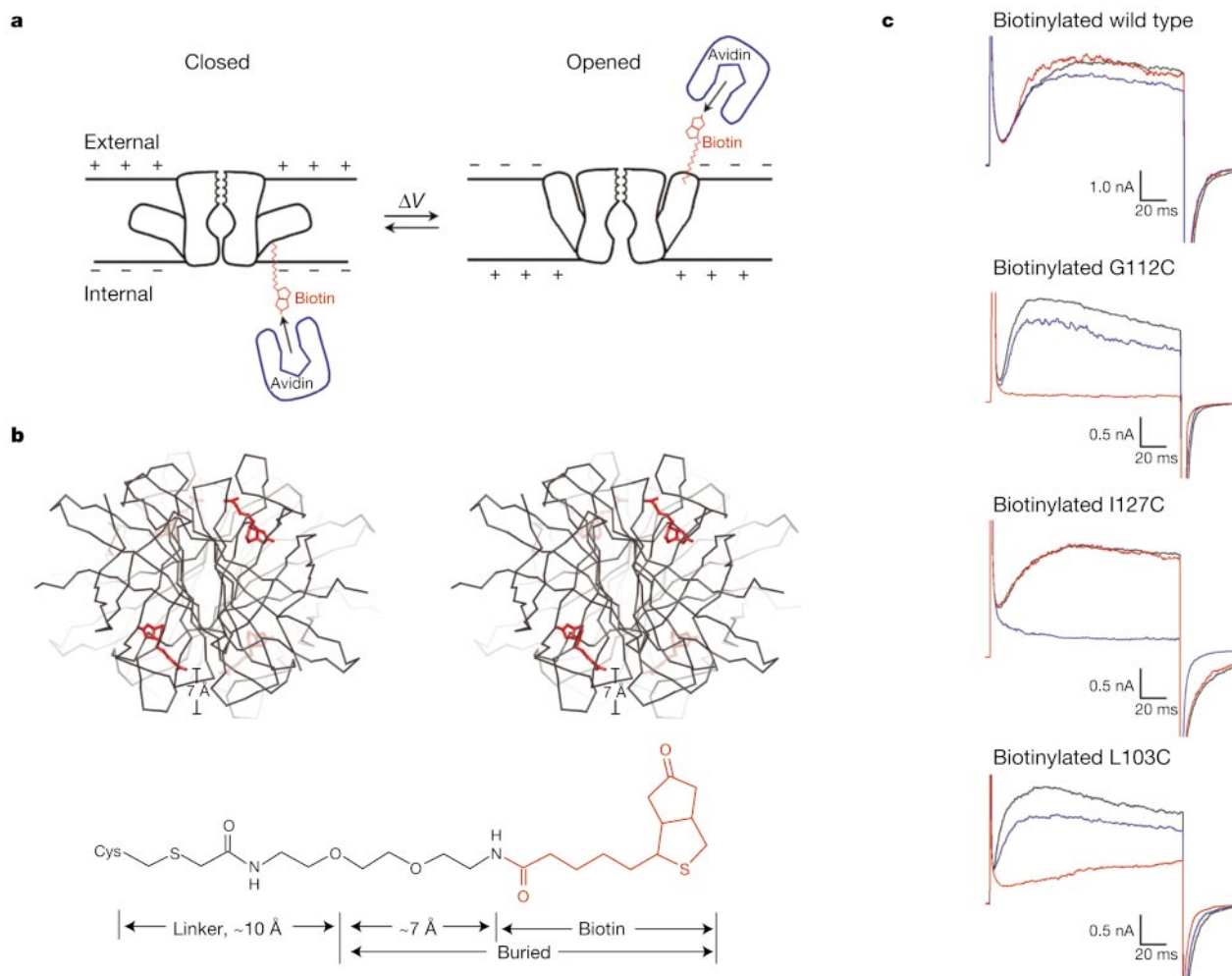


Figure 2 Using avidin and tethered biotin as a molecular ruler to measure positions of the voltage-sensor paddles. **a**, Experimental strategy: KvAP channels with biotin tethered to a site-directed cysteine can be 'grabbed' by avidin in solution to affect channel function. **b**, Stereo view of an avidin tetramer (α trace, Protein Data Bank code 1AVD) with biotin (red) in its binding pockets. Chemical structure of biotin and its PEO-iodoacetyl linker with

buried (inside avidin) and exposed segments indicated. **c**, Representative traces showing the effects of avidin on wild-type and mutant biotinylated KvAP channels. Currents were elicited with depolarizing steps in the absence (black traces) or presence of internal (blue traces) or external (red traces) avidin.

ascribed to two structurally distinct populations of channels. We conclude that positions 121 and 122 actually drag biotin and its linker all the way across the lipid membrane from the solution on one side to that on the other when the channel gates. This finding is very important, because it indicates that the voltage-sensor paddles must move a large distance through the membrane, and that they must move through a lipid environment where a bulky chemical structure such as biotin and its linker (Fig. 2b) would be unimpeded. Biotin and its linker could not be dragged through the core of a protein, as would be required by conventional models, which invoke an S4 helix buried within the protein.

Discussion

Positional constraints on the voltage-sensor paddles are summarized in Fig. 5a, b. Horizontal solid lines show the external and internal surfaces of the cell membrane (~ 35 Å thick) and dashed lines show the 10 Å limit from the membrane surface set by the linker length (Fig. 2b). If the α -carbon of a cysteine residue comes within 10 Å of the membrane surface, then avidin can bind to the attached biotin, otherwise the biotin is inaccessible. Avidin is too

large to enter crevices on the channel, so it cannot penetrate below the membrane surface.

At negative membrane voltages when the channel is closed, no positions are accessible to the external side; all residues on the voltage-sensor paddles in their channel-closed position must lie deeper in the membrane than the 10-Å limit below the external surface (Fig. 5a). At negative voltages, the blue and yellow residues on S4 bind from inside, and therefore must come within 10 Å of the internal solution. The red residues, including all of S3b, the tip of the paddle and the first helical turn into S4, are protected from both sides at negative voltages and must therefore lie further than 10 Å from both surfaces; that is, between the two dashed lines. This pattern of accessibility in the closed (negative voltage) conformation constrains the voltage-sensor paddles to lie near the internal surface of the membrane with S3b above S4, as shown (Fig. 5a), similar to the orientation in the crystal structure⁴.

At positive membrane voltage when the channel is opened, the entire S3b helix, the tip of the paddle and the first two-and-a-half helical turns of S4 become accessible to external avidin and must therefore be within 10 Å of the external solution (Fig. 5b). The next

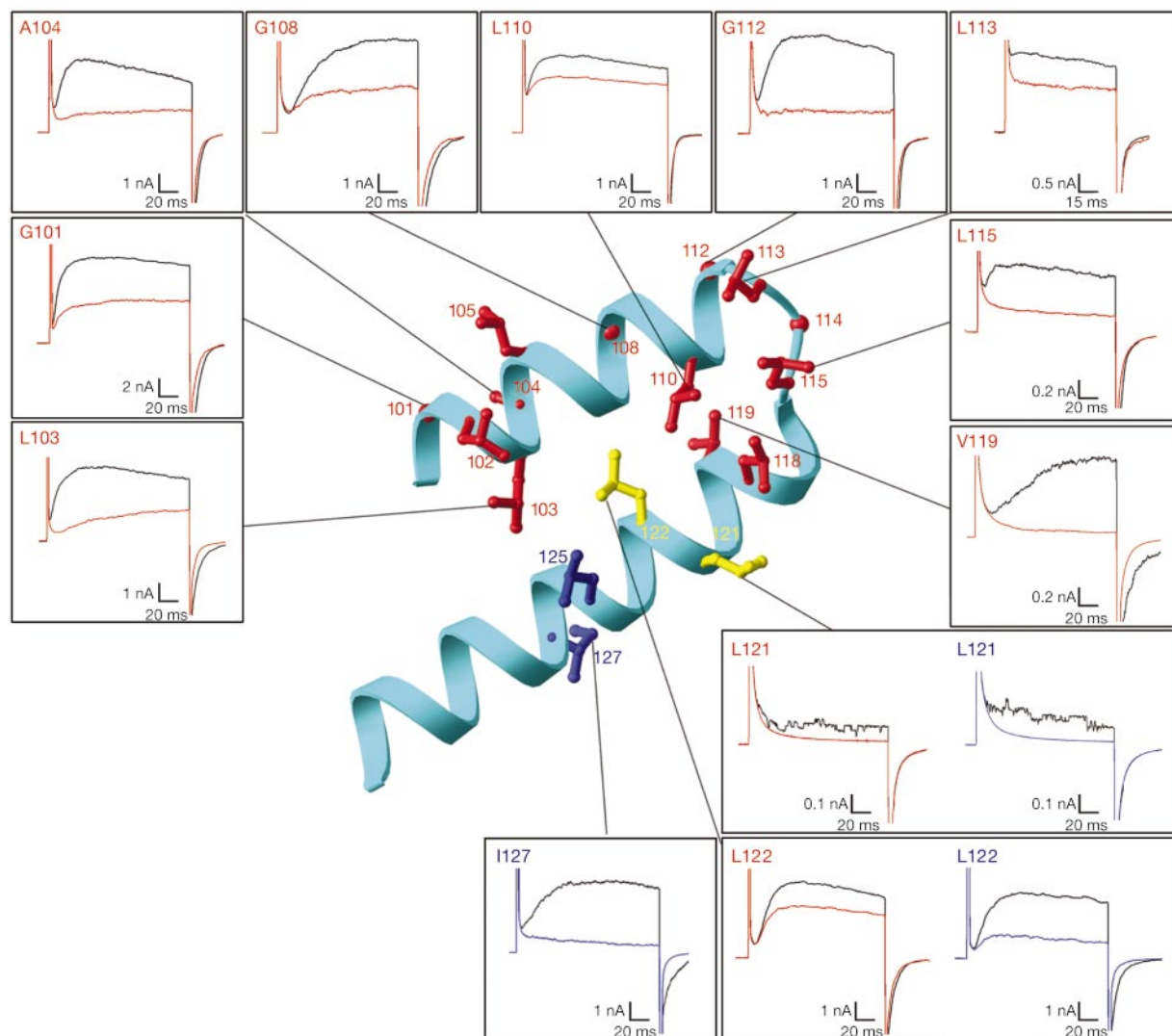


Figure 3 Accessibility of voltage-sensor paddle residues to the internal and external sides of the membrane. Membrane was depolarized every 120 s in the absence (control) or presence of avidin on the internal and external side of each biotinylated mutant. Red side chains on the voltage-sensor paddle and selected traces indicate inhibition by avidin from the external side only; blue side chains and selected traces indicate inhibition by avidin

from the internal side only; yellow side chains and blue and red traces indicate inhibition by avidin from both sides. Black traces show control currents before avidin addition and coloured traces show currents after adding avidin. Each trace is the average of 5–10 measured traces, with the exception of the single traces for position 121.

helical turn on S4 (positions 125 and 127, blue residues) remains more distant than 10 Å from the external surface. The Fabs inform us that, in the opened conformation, two helical turns of S3b and one turn of S4 must actually protrude clear into the external solution, otherwise the epitope would not be exposed (Fig. 1b). The Fabs and pattern of avidin accessibility in the opened (positive voltage) conformation constrain the voltage-sensor paddles to be near the external membrane surface with a more vertical orientation, as shown (Fig. 5b).

In moving from their closed to their opened position, the voltage-sensor paddles' centre of mass translates approximately 20 Å (assuming a membrane thickness of 35 Å) through the membrane from inside to out, and the paddles tilt from a somewhat horizontal to a more vertical orientation. Arginines 117, 120, 123 and 126, the first four arginines on S4 (Fig. 1a), are distributed along the paddles. In the closed conformation, arginines 126, 123 and perhaps 120 can probably extend their positively charged guanidinium group to the internal lipid head group layer, whereas arginine 117 is near the internal side but still within the membrane. In the opened conformation, arginines 117, 120 and probably 123 can extend to the external solution or lipid head group layer, whereas 126 is near the external side but still within the membrane. The near-complete transfer of these four arginine residues across the membrane (in each of the four subunits) is compatible with the total gating charge in the *Shaker* K⁺ channel of 12–14 electrons (3.0–3.5 electrons per subunit)^{15–17}, and with the demonstration that each of these four arginine residues carries approximately one electron charge unit^{16,17}.

A positional aspect of the voltage-sensor paddles not constrained by these experiments is whether they lie tangential to the channel's outer surface or whether they point in a radial direction away from it. The flexible S3 loops and S4–S5 linkers probably do not constrain the paddles much. However, we have two reasons for thinking that the paddles are positioned tangentially, which is the way they are positioned in Fig. 5c. First, the paddles are oriented tangentially in the crystal structure of the full channel; and second, the crystal structure of the isolated voltage sensor shows interesting salt-bridge interactions between S4 and S2, and between S3a and S2 (ref. 4). A tangential orientation would favour salt-bridge interactions between arginine residues on the voltage-sensor paddles and acidic residues on the S2 and S3a helices. Studies by Papazian and co-workers on the *Shaker* K⁺ channel indeed suggest that salt-bridge pairs are important, and that they might exchange as the paddles

move between their closed and opened positions on the outer perimeter of the channel^{18–20}.

To a first approximation, we describe the voltage-sensor paddles as hydrophobic cations that carry gating charges through the lipid bilayer. Ionic interactions between S4 arginines and S2 and S3 acidic residues probably assist the gating charge movement, and the presence of a polar, sometimes acidic loop between S3b and S4 in certain voltage-dependent K⁺ channels (for example, the *Shaker* K⁺ channel²¹) raises interesting questions about the structure of the lipid–water interface above the paddles when they are in their closed channel position. However, there is no escaping the basic finding that the paddles are located at the protein–lipid interface, and move while contacting the lipid membrane. Given that the paddles move through a lipid environment, it is interesting to ask why the basic residues on the voltage-sensor paddles are nearly always arginine and not lysine? One reason is that arginine ($pK_a \approx 12.5$ in water) will nearly always move through the membrane with a protonated, charged side chain, whereas lysine ($pK_a \approx 10.5$ in water) will sometimes be unprotonated. Another reason may be that arginine can easily participate in multiple hydrogen bonds (perhaps with acidic side chains) in a spatially directed way. Yet a third possible reason is that it might be energetically less costly to transfer from water to lipid the diffuse positive charge on a guanidinium group (arginine) than the more focused charge on an amino group (lysine). Studies of the transfer of hydrophobic peptides from water to octanol by White and co-workers²² provide evidence for this idea by showing that the charged lysine side chain is energetically more costly in this assay than the charged arginine by about 1.0 kcal mol^{−1}. Given that four voltage-sensor paddles must move through the membrane to open the pore, we might expect there to be a strong evolutionary bias in favour of arginine over lysine. In this regard, it is interesting that the recent structure of MscS, a mechanosensitive channel, has two basic amino acids in its transmembrane segments that are arginine²³. MscS is not gated by voltage *per se*, but its mechanical force-induced opening can be modulated by voltage²⁴, and the candidate residues that are proposed to underlie voltage modulation are arginine, not lysine²³.

On the basis of the KvAP crystal structure⁴, the deduced positions of its voltage-sensor paddles in the functioning channel (Fig. 5a, b), and previous studies of opened and closed K⁺ channels^{25,26}, we propose a model for how membrane voltage gates the pore (Fig. 5c, d). This is a working model to envision how the electromechanical coupling process might occur, and it will need to be revised as more data are obtained. In the closed conformation (Fig. 5c), the positively charged voltage-sensor paddles (red) are near the intracellular membrane surface, held there by the large electric field (mean value $>10^7$ V m^{−1}) imposed by the negative resting membrane voltage. In this conformation, the S5 and S6 (outer and inner) helices are arranged as they are in KcsA, a closed K⁺ channel²⁵; favourable packing interactions between the inner helices have been proposed to help to stabilize the closed conformation²⁷. In response to depolarization, the voltage-sensor paddles move across the membrane to their external position, which exerts a force on the S4–S5 linker, pulling the S5 helices away from the pore axis (Fig. 5d). The crystal structure of KvAP shows that the S5 helices form a cuff outside the S6 helices, and suggests that when the S5 cuff is expanded, the S6 helices follow, opening the pore⁴.

Although previous mutational studies of voltage-dependent gating have been interpreted in the context of the conventional models, many of the data are consistent with the structural model presented here, and indeed help to constrain it. For example, second-site suppressor mutations in the *Shaker* K⁺ channel suggest that certain salt bridges probably break and reform as the voltage-sensor paddles move²⁰. In addition to four S4 arginine residues, a component of the gating charge in *Shaker* was shown to come from an S2 acidic residue (glutamate 293, corresponding to aspartate 72 in KvAP)¹⁶, implying that S2 might change its position in the membrane as the

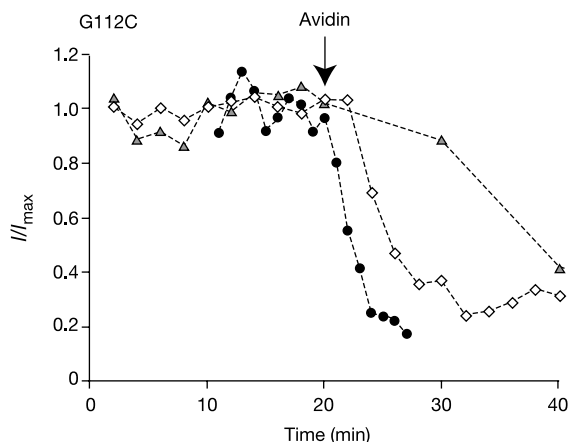


Figure 4 Exposure of the voltage-sensor paddle to the external solution occurs only when the membrane is depolarized. For the biotinylated G112C mutant, normalized (to the average control) currents elicited by depolarization to 100 mV every 60 s (circles), 120 s (diamonds) and 600 s (triangles) are shown before and after the addition of avidin to the external side.

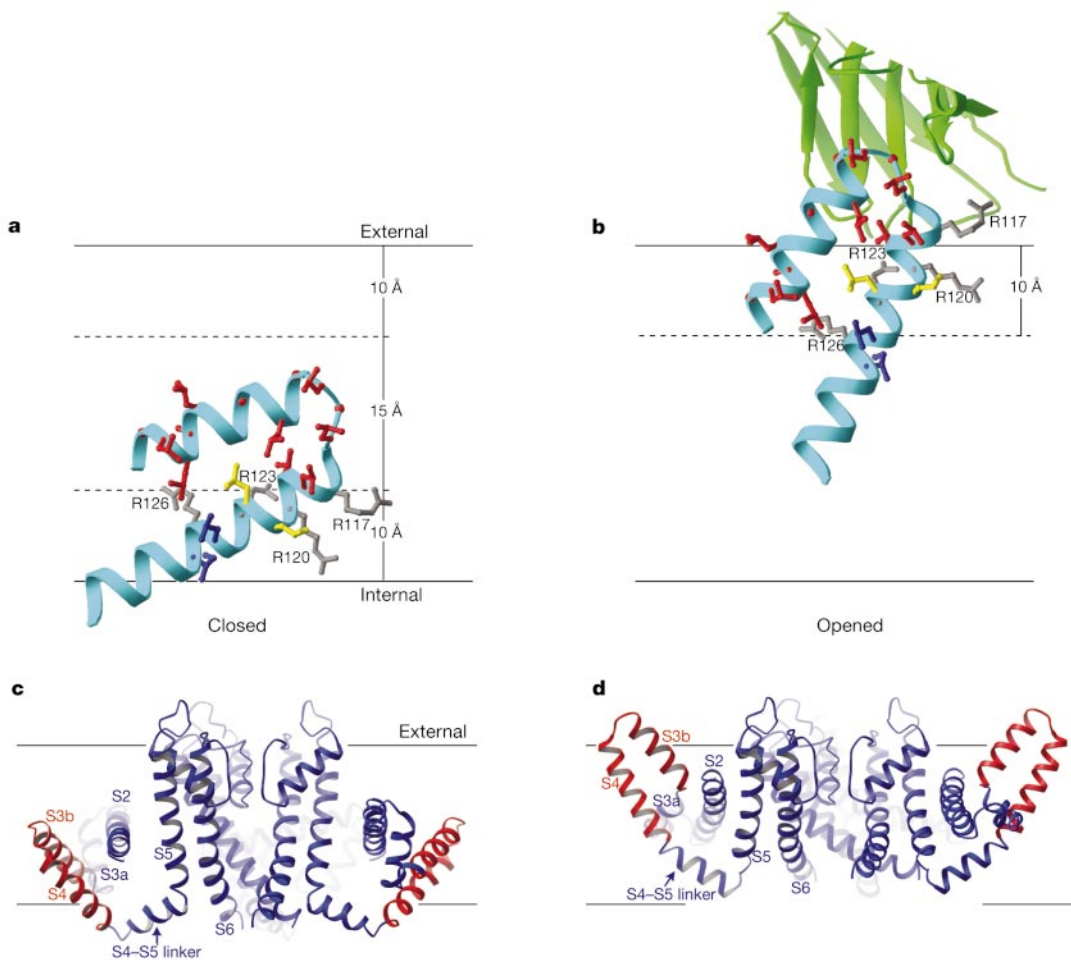


Figure 5 Positions within the membrane of the voltage-sensor paddles during closed and opened conformations, and a hypothesis for coupling to pore opening. **a**, **b**, Closed (**a**) and opened (**b**) positions of the paddles derived from the tethered biotin-avidin measurements, and structural and functional measurements with Fabs. A voltage-sensor paddle is shown as a cyan ribbon with side chains colour-coded as in Fig. 3. Grey side chains show four arginine residues on the paddle, and the green ribbon (**b**) shows part of a

bound Fab from the crystal structures. Solid horizontal lines show the external and internal membrane surfaces, and dashed lines indicate the 10-Å distance from the surfaces set by biotin and its linker. **c**, The closed KvAP structure is based on the paddle depth and orientation in **a** (red), and adjusting the S5 and S6 helices of KvAP to the positions in KcsA, a closed K⁺ channel. **d**, The opened KvAP structure is based on the paddle depth and orientation in **b** (red), and the pore of KvAP.

voltage-sensor paddles move; the loose attachment of S2 to the pore in the KvAP crystal structure certainly would allow this to happen. Disulphide cross-linking of S4 to the turret loop between S5 and the pore helix²⁸, and of two S4 segments within a tetramer to each other²⁹ (an observation that was very difficult to understand in the context of conventional models), are in agreement with the mobile voltage-sensor paddles in Fig. 5c, d. The accessibility of thiol-reactive compounds to cysteine residues introduced into the *Shaker* K⁺ channel S4 (refs 30–32), and the ability of histidine residues to shuttle protons across the membrane³³, are in reasonable agreement with our data on voltage-sensor paddle movements. However, our structural and mechanistic interpretation, summarized in Fig. 5, differs fundamentally from past models of voltage-dependent gating. This new picture is based on the elucidation of the voltage-sensor paddle structure, its flexible attachments and disposition relative to the pore, and the paddles' positions in the membrane when the channel is closed and opened.

Conclusion

Here and in an accompanying paper⁴, we have shown that (1) the gating charges are carried on voltage-sensor paddles, which are helix–turn–helix structures attached to the channel through flexible

S3 loops and S4–S5 linkers; (2) the paddles are located at the channel's outer perimeter and move within the lipid membrane; (3) the S2 helices lie beside the pore and contain acidic amino acids that could help to stabilize positive charges on the paddles as they move across the membrane; (4) the total displacement of the voltage-sensor paddles is approximately 20 Å perpendicular to the membrane; and (5) the large displacement of the paddles could open the pore by pulling on the S4–S5 linker. We conclude that the voltage sensor operates by an extraordinarily simple principle based on hydrophobic cations attached to levers, which enables the membrane electric field to perform mechanical work to open and close the ion-conduction pore. □

Methods

Biotinylation

All biotinylation studies were carried out using a KvAP channel in which the single endogenous cysteine was mutated to serine (C247S). This mutant showed no detectable electrophysiological differences when compared to wild-type KvAP. Single cysteine mutations were then added to the voltage-sensor paddle (positions 101 to 127) using the QuickChange method (Stratagene) and confirmed by sequencing the entire gene. Mutant channels were expressed and purified by the same protocol as wild-type KvAP channels⁵, except that before gel filtration, mutant KvAP channels were incubated with 10 mM DTT for 1 h. Immediately after gel filtration, mutant channels (at 0.5–1.0 mg ml⁻¹) were

incubated with 500 μM PEO-iodoacetylbiotin (Pierce) for 2–3 h at room temperature, and then either reconstituted into lipid vesicles for electrophysiological analysis, or purified away from the excess biotin reagent on a desalting column, complexed with avidin (Pierce) and run on an SDS gel to assess the extent of biotinylation.

Electrophysiology

Fabs (6E1 and 33H1) and VSTX1 were purified as described in the companion paper⁴ and used in electrophysiological assays. Electrophysiological studies of wild-type and biotinylated KvAP channels were carried out as described⁵. To measure inhibition, Fabs (~500 nM), VSTX1 (30 nM) or avidin (40 $\mu\text{g ml}^{-1}$) were added to wild-type and/or cysteine-mutant, biotinylated channels reconstituted into lipid membranes, and studied using various voltage protocols.

Received 19 February; accepted 11 March 2003; doi:10.1038/nature01581.

- Sigworth, F. J. Voltage gating of ion channels. *Q. Rev. Biophys.* **27**, 1–40 (1994).
- Armstrong, C. M. & Bezanilla, F. Charge movement associated with the opening and closing of the activation gates of the Na^+ channels. *J. Gen. Physiol.* **63**, 533–552 (1974).
- Bezanilla, F. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* **80**, 555–592 (2000).
- Jiang, Y. *et al.* X-ray structure of a voltage-dependent K^+ channel. *Nature* **423**, 33–41 (2003).
- Ruta, V., Jiang, Y., Lee, A., Chen, J. & MacKinnon, R. Functional analysis of an archaebacterial voltage-dependent K^+ channel. *Nature* **422**, 180–185; advance online publication, 2 March 2003 (doi:10.1038/nature01473).
- Swartz, K. J. & MacKinnon, R. Mapping the receptor site for hanatoxin, a gating modifier of voltage-dependent K^+ channels. *Neuron* **18**, 675–682 (1997).
- Slatin, S. L., Qiu, X. Q., Jakes, K. S. & Finkelstein, A. Identification of a translocated protein segment in a voltage-dependent channel. *Nature* **371**, 158–161 (1994).
- Qiu, X. Q., Jakes, K. S., Finkelstein, A. & Slatin, S. L. Site-specific biotinylation of colicin Ia. A probe for protein conformation in the membrane. *J. Biol. Chem.* **269**, 7483–7488 (1994).
- Qiu, X. Q., Jakes, K. S., Kienker, P. K., Finkelstein, A. & Slatin, S. L. Major transmembrane movement associated with colicin Ia channel gating. *J. Gen. Physiol.* **107**, 313–328 (1996).
- Pugliese, L., Coda, A., Malcovati, M. & Bolognesi, M. Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7 Å resolution. *J. Mol. Biol.* **231**, 698–710 (1993).
- Zagotta, W. N., Hoshi, T., Dittman, J. & Aldrich, R. W. Shaker potassium channel gating. II. Transitions in the activation pathway. *J. Gen. Physiol.* **103**, 279–319 (1994).
- Zagotta, W. N., Hoshi, T. & Aldrich, R. W. Shaker potassium channel gating. III. Evaluation of kinetic models for activation. *J. Gen. Physiol.* **103**, 321–362 (1994).
- Schoppa, N. E. & Sigworth, F. J. Activation of Shaker potassium channels. III. An activation gating model for wild-type and V2 mutant channels. *J. Gen. Physiol.* **111**, 313–342 (1998).
- Cole, K. S. & Moore, J. W. Potassium ion current in the squid giant axon: dynamic characteristic. *Biophys. J.* **1**, 1–14 (1960).
- Schoppa, N. E., McCormack, K., Tanouye, M. A. & Sigworth, F. J. The size of gating charge in wild-type and mutant Shaker potassium channels. *Science* **255**, 1712–1715 (1992).
- Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K^+ channel. *Neuron* **16**, 1159–1167 (1996).
- Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K^+ channel. *Neuron* **16**, 1169–1177 (1996).
- Papazian, D. M. *et al.* Electrostatic interactions of S4 voltage sensor in Shaker K^+ channel. *Neuron* **14**, 1293–1301 (1995).
- Tiwari-Woodruff, S. K., Lin, M. A., Schulteis, C. T. & Papazian, D. M. Voltage-dependent structural interactions in the Shaker K^+ channel. *J. Gen. Physiol.* **115**, 123–138 (2000).
- Papazian, D. M., Silverman, W. R., Lin, M. C., Tiwari-Woodruff, S. K. & Tang, C. Y. Structural organization of the voltage sensor in voltage-dependent potassium channels. *Novartis Found. Symp.* **245**, 178–190 (2002).
- Gonzalez, C., Rosenman, E., Bezanilla, F., Alvarez, O. & Latorre, R. Periodic perturbations in Shaker K^+ channel gating kinetics by deletions in the S3–S4 linker. *Proc. Natl Acad. Sci. USA* **98**, 9617–9623 (2001).
- Wimley, W. C., Creamer, T. P. & White, S. H. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109–5124 (1996).
- Bass, R. B., Strop, P., Barclay, M. & Rees, D. C. Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science* **298**, 1582–1587 (2002).
- Martina, B., Buechner, M., Delcour, A. H., Adler, J. & Kung, C. Pressure-sensitive ion channel in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **84**, 2297–2301 (1987).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
- Yifrach, O. & MacKinnon, R. Energetics of pore opening in a voltage-gated K^+ channel. *Cell* **111**, 231–239 (2002).
- Laine, M. *et al.* Structural interactions between voltage sensor and pore in the Shaker K^+ channels. *Biophys. J.* **82**, 231a (2002).
- Aziz, Q. H., Partridge, C. J., Munsey, T. S. & Sivaprasadarao, A. Depolarization induces intersubunit cross-linking in a S4 cysteine mutant of the Shaker potassium channel. *J. Biol. Chem.* **277**, 42719–42725 (2002).
- Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the Shaker K^+ channel S4. *Neuron* **16**, 387–397 (1996).
- Yusaf, S. P., Wray, D. & Sivaprasadarao, A. Measurement of the movement of the S4 segment during the activation of a voltage-gated potassium channel. *Pflügers Arch.* **433**, 91–97 (1996).
- Baker, O. S., Larsson, H. P., Mannuzzu, L. M. & Isacoff, E. Y. Three transmembrane conformations and sequence-dependent displacement of the S4 domain in Shaker K^+ channel gating. *Neuron* **20**, 1283–1294 (1998).
- Starace, D. M., Stefani, E. & Bezanilla, F. Voltage-dependent proton transport by the voltage sensor of the Shaker K^+ channel. *Neuron* **19**, 1319–1327 (1997).

Acknowledgements We thank D. Gadsby and O. Andersen for helpful discussions and advice on the manuscript. This work was supported in part by a grant from the National Institutes of Health (NIH) to R.M. V.R. is supported by a National Science Foundation Graduate Student Research Fellowship, and R.M. is an Investigator in the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M. (mackinn@rockvax.rockefeller.edu).

Ligand–receptor binding revealed by the TNF family member TALL-1

Yingfang Liu^{*†}, Xia Hong^{*†}, John Kappler^{*‡§}, Ling Jiang^{*}, Rongguang Zhang^{||}, Liangguo Xu^{*}, Cheol-Ho Pan^{*}, Wesley E. Martin^{*}, Robert C. Murphy^{*}, Hong-Bing Shu^{*¶}, Shaodong Dai^{*‡} & Gongyi Zhang^{*§}

^{*} Integrated Department of Immunology, National Jewish Medical and Research Center, [‡] Howard Hughes Medical Institute, and [§] Department of Pharmacology, Biomolecular Structure Program, School of Medicine, University of Colorado Health Science Center, 1400 Jackson Street, Denver, Colorado 80206, USA

[¶] Department of Cell Biology and Genetics, College of Life Sciences, Peking University, Beijing 100871, China

^{||} Structural Biology Section, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA

[†] These authors contributed equally to this work

The tumour necrosis factor (TNF) ligand TALL-1 and its cognate receptors, BCMA, TACI and BAFF-R, were recently identified as members of the TNF superfamily, which are essential factors contributing to B-cell maturation. The functional, soluble fragment of TALL-1 (sTALL-1) forms a virus-like assembly for its proper function. Here we determine the crystal structures of sTALL-1 complexed with the extracellular domains of BCMA and BAFF-R at 2.6 and 2.5 Å, respectively. The single cysteine-rich domain of BCMA and BAFF-R both have saddle-like architectures, which sit on the horseback-like surface formed by four coil regions on each individual sTALL-1 monomer. Three novel structural modules, D2, X2 and N, were revealed from the current structures. Sequence alignments, structural modelling and mutagenesis revealed that one disulphide bridge in BAFF-R is critical for determining the binding specificity of the extracellular domain eBAFF-R to TALL-1 instead of APRIL, a closely related ligand of TALL-1, which was confirmed by binding experiments *in vitro*.

TNF family ligands and their corresponding receptors (TNF-Rs) have pivotal roles in many biological processes in mammalian cells, such as in host defence, inflammation, apoptosis, autoimmunity and organogenesis. At least 18 TNF ligands and 28 receptors have been identified so far. Some ligands have multiple receptors, and some receptors also bind multiple ligands. The interactions between ligands and receptors are usually very specific and have high affinity (0.1–1 nM)¹.

Ligands such as TALL-1/BAFF/THANK/BLyS/zTNF4 and APRIL/TALL-2, as well as receptors such as BCMA, TACI and BAFF-R/BR3, are recently identified members of the TNF/TNF-R family^{2–15}. Overexpression of sTALL-1 in mice leads to increased numbers of mature B lymphocytes, splenomegaly, anti-DNA antibodies, proteinuria and glomerulonephritis. These phenotypes mimic those of SLE systemic lupus erythematosus^{7–13}. TALL-1 deficiency in mice leads to the complete arrest of peripheral B cells at the T1 stage^{16,17}. The phenotype is similar to that caused by BAFF-R deficiency^{14,15}. The knockout of BCMA does not lead to any severe B-cell phenotypes¹⁸. It therefore seems that BCMA is dispensable for B-cell maturation. Knockout of TACI increased the total number of circulating B cells. However, the ratio of maturing B-cell subsets in the spleen was normal¹⁹. The role of TACI in the B-cell maturation process is still a mystery.

APRIL/TALL-2, the closest family member of TALL-1, has a low abundance in normal tissues but is present at high concentrations in transformed cell lines and in a variety of human cancers²⁰. Recent data show that BAFF-R does not bind APRIL^{14,15}, suggesting that APRIL is dispensable for B-cell maturation^{14,15,21}. Nevertheless, APRIL-deficient mice die *in utero*²². The role of APRIL is still unknown.

All other known structures of TNF ligands exist as trimers made entirely of β -strands and loops with a standard ‘jellyroll’ topology^{23–28}. This is also true for sTALL-1 at low pH^{29,30}. At pH 7.4 or higher, sTALL-1 forms a virus-like cluster through an unusual ‘flap’ region on sTALL-1 (ref. 31). Transfection and B-cell stimulation assays showed that this oligomerization status is required for its proper function³¹.

The structure of the complex of TNF- β with cysteine-rich

domains (CRDs) from its cognate receptor TNF-R1 has been determined³². This structure shows that the three elongated receptor domains bind to one TNF trimer at the interfaces formed between the TNF monomers³². Two CRDs (CRD2 and CRD3) make contacts with two distinct regions of TNF- β . The recently determined complex structure of TRAIL and DR5 disclosed a similar interaction mode to that observed in the TNF- β and TNF-R1 co-crystal structure, although CRD3 of DR5 assumes a different orientation from that in the TNF- β and TNF-R1 structures^{33,34}.

In contrast to the other receptor family members that have at least three or four CRDs in their extracellular domains, BAFF-R has only a half CRD (one module), BCMA has only one CRD, and TACI has two CRDs^{6–9,14,15}. Nevertheless, the overall binding affinities of sTALL-1 with BCMA, BAFF-R and TACI are similar to those of other family members (0.1–1 nM)^{31,35}. Furthermore, as predicted from sequence alignment, the CRDs in BCMA and TACI contain A1 and C2 modules⁷, which are two of the multiply defined structural motifs that characterize the extracellular domains of TNF receptors³⁶. The C2 module was also found in TNF-R1 and Fn14 (ref. 37). However, the C2 in TNF-R1 is not involved in ligand binding³⁶. The only half CRD in BAFF-R was originally predicted to be the C2 module^{14,15}. However, this CRD has recently been named as a new module, X2 (ref. 37). It is likely that there are novel interactions between these unique ligand–receptor couples that account for their high affinity^{31,37}. In testing this hypothesis, we determine here the structures of sTALL-1 complexed with the extracellular domains of either BCMA (eBCMA) or BAFF-R (eBAFF-R).

Structural determination

Crystals of sTALL-1 with eBCMA and eBAFF-R were obtained by diffusing the receptor fragments into the sTALL-1 crystals (see Methods). The structures of both complexes were determined by difference Fourier analysis with the use of the available sTALL-1 model (Fig. 1a, Table 1 and Supplementary Information). The structure of sTALL-1 with eBCMA has been refined to an *R*-factor of 20.9% (*R*_{free}-factor 23.4%) against data to 2.6 Å resolution in space group P6₃22, with 10 sTALL-1 monomers and seven complete molecules and one partial eBCMA molecule in the asymmetric unit

(unit cell $234 \times 234 \times 217 \text{ \AA}$) (Table 1). Owing to crystal packing, another two receptor-binding sites were left unoccupied. Similar results are seen for the complex between sTALL-1 and eBAFF-R, with a final resolution of $2.5 \text{ \AA}$ (Table 1), although the corresponding partial eBAFF-R is almost completely disordered. The current model of the eBCMA monomer contains residues 5–43 (Fig. 1b). The model of eBAFF-R contains residues 16–45 (Fig. 1c).

Overall structure

The space group of the sTALL-1 crystals remained $P6_322$ with the same cell dimensions with or without receptors bound. There are two virus-like clusters in one unit cell. Each cluster has 60 copies of sTALL-1, 42 fully occupied eBCMA or eBAFF-R, and 6 partial copies of eBCMA or eBAFF-R. There are 12 copies of sTALL-1 free of receptors owing to crystal packing. All receptors are located on the outer extreme shell, which expands the ball-like shell another $\sim 20 \text{ \AA}$ in each direction. The overall arrangement of the receptors on the shell looks like a sunflower, with receptors as the flower petals and sTALL-1 as the seedbed (Fig. 2). Molecules shown in red in Fig. 2 are missing from the complex structure; molecules shown in blue are partly occupied. The conformational change in sTALL-1 is negligible before and after receptor binding, which is the only similarity between this interaction and that of other TNF family members.

Four lines of evidence support the proposal that our crystal structure reflects the actual interactions of the complexes in solution and *in vivo*. First, coexpression of sTALL-1 with eBCMA or eBAFF-R generates the virus-like cluster in solution as judged by gel-filtration studies and SDS–polyacrylamide gel electrophoresis analysis at an approximate sTALL-1:eBCMA or sTALL-1:eBAFF-R

ratio of 1:1. Different salt concentrations (from 100 mM to 1 M NaCl) produce the same elution profile, in which complexes of sTALL-1 with eBCMA or eBAFF-R are eluted at the void volume from a Superdex-200 column. Thus, binding between ligand and receptors is stable and insensitive to salt concentration. Second, in receptor-soaked sTALL-1 crystals, all seven fully occupied receptors and one partial receptor have equivalent binding sites on sTALL-1 in the asymmetric unit, so the binding is highly specific. Third, eBCMA and eBAFF-R have similar binding modes and occupy the same site on sTALL-1. Last, each of the three carboxy termini of eBCMA and eBAFF-R on the sTALL-1 trimer point in the same direction—towards the putative membrane surface for trimerization. Thus, we believe that the interactions revealed from the complex structures represent TALL-1–BCMA and TALL-1–BAFF-R *in vivo*.

Structure of eBCMA

As predicted, eBCMA contains two modules. One is the A1 module, consisting of three β -strands, with strands 1 and 3 linked by the only disulphide bridge³⁶. The other module is C2-like (the two disulphide bridges formed are Cys III–Cys VI and Cys IV–Cys V), but the disulphide arrangement is the same as in a typical B2 module (the two disulphide bridges formed are Cys III–Cys V and Cys IV–Cys VI)³⁶ (Fig. 1b). For clarity, we temporarily termed this module D2 because of its difference from C2 and B2. There are two short helices in the D2 module that are located at the amino terminus and the C terminus of the module; one is from Cys III to Cys IV and the other is from Cys V to Cys VI. The latter helix extends further after the disulphide bridge and forms a helix three turns long, which is unique among all known TNF receptor structures. The arrangement of modules A1 and D2 of eBCMA is similar to that of A1 and C2 in

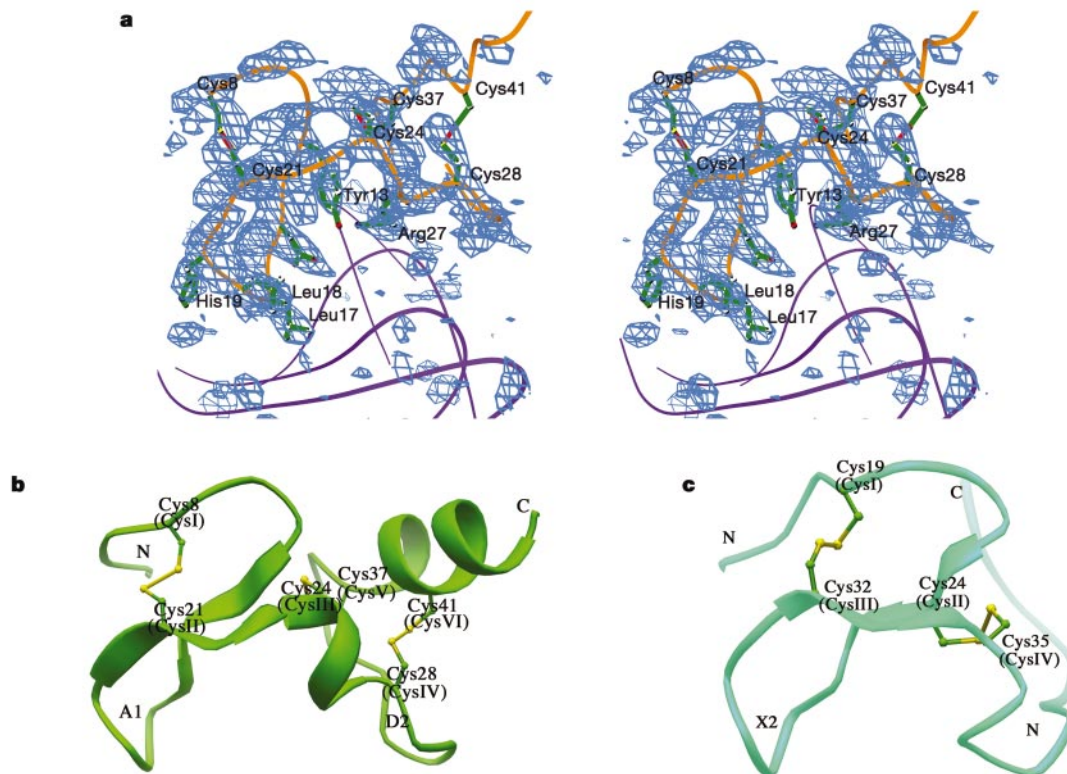


Figure 1 Structures of eBCMA and eBAFF-R. **a**, Initial $F_0 - F_c$ map of eBCMA with sTALL-1 at the 2.5σ level. Phases are calculated from the sTALL-1 model (Protein Data Bank ID 1JH5). eBCMA is shown as the final refined model. The portion of the map shown is representative of all eight bound receptors in the asymmetric unit. Most residues are shown with their side chains. **b**, Ribbon diagram of the three-dimensional structure of

eBCMA (residues 5–43), showing three disulphide bridges. **c**, Ribbon diagram of the three-dimensional structure of eBAFF-R (residues 16–45), showing two disulphide bridges. All figures were prepared with RIBBONS⁴⁴ except **a**, which was prepared with BOBSCRIPT⁴⁵.

Table 1 Experimental data on crystal structure determination and refinement

Data set	Resolution (Å)	R_{merge} (%)	No. of unique reflections	Total observations	Completeness (%)	R -factor	R_{free}	$\langle I \rangle / \langle \sigma(I) \rangle$
eBCMA	2.6	12.9 (55.0)	97,672	1,056,950	94.0	20.9 (33.8)	23.5 (36.0)	10.9
eBAFF-R	2.5	13.5 (60.0)	121,940	1,415,809	99.3	24.5 (46.0)	26.5 (43.7)	14.4

$R_{\text{merge}} = \sum |I_j - \langle I \rangle| / \sum I_j$ with Bijvoet pairs treated as equivalent. Total observations are the number of full and partial observations measured with non-negative intensity to the indicated resolution. Completeness is the percentage of possible unique reflections measured with $I/\sigma(I) \geq -3$ (default value of SCALEPACK⁴²) to the indicated resolution. R -factor = $\sum |F_o - F_c| / \sum F_o$ for all amplitudes with $F/\sigma(F) \geq 0$ ($I/\sigma(I) \geq -3$) measured; R_{free} is calculated with 5% of the data. Data in parentheses are R values for the highest-resolution bins. $\langle I \rangle / \langle \sigma(I) \rangle$ is the ratio of the average value of I to $\sigma(I)$. There is a total of 13,743 atoms in the final refinement for the eBCMA and sTALL-1 complex, and there is one water molecule, with r.m.s.d. bonds = 0.008 and r.m.s.d. angles = 1.58. There is a total of 13,230 atoms in the final refinement for the eBAFF-R and sTALL-1 complex, with r.m.s.d. bonds = 0.008 and r.m.s.d. angles = 1.7.

TNF-R1 (ref. 36). Modules A1 and D2 form a saddle-like structure with each module as half of the saddle and the unique helix as the 'rider' (Fig. 1b). From the initial $2F_o - F_c$ and $F_o - F_c$ maps, all seven copies of the A1 modules are very rigid, with temperature factors similar to that of sTALL-1, and most side chains are ordered. The partial copy of the eighth eBCMA contains only the A1 module. The root-mean-square deviation (r.m.s.d.) is 0.2 Å for eight A1 modules in the asymmetry unit. In contrast, the D2 modules are relatively flexible, especially in the region between Cys IV and Cys V. The r.m.s.d. is 1.5 Å for the seven D2 modules in the asymmetry unit. The eighth D2 is severely disordered except for the region between Cys III and Cys IV.

Structure of eBAFF-R

To our surprise, most of eBAFF-R has a similar folding to that of eBCMA, although it was predicted that eBAFF-R contains only one C2 or X2 module^{14,15,37} (Fig. 1c). The structure of eBAFF-R shows that it contains one major module, X2, and one minor N-shaped module (with no disulphide bridge), which we temporarily named module N. The X2 module contains four cysteines, which form two disulphide bridges. The X2 module is similar to the A1 model in eBCMA, but the additional disulphide bridge is typical of the B2 instead of the A2 module. We temporarily named this module X2, following the name from the literature³⁷. For the N module, there is a one-turn helix and a coil eight residues long. Interestingly, from the initial $2F_o - F_c$ and $F_o - F_c$ maps, the coil region is well ordered, although it does not seem to make any contact with the other part of eBAFF-R. We believe that the three proline residues in the coil region account for this rigidity. We conclude that BAFF-R has the smallest functional motif in its extracellular part among all

known TNF receptor members. A partial model of BAFF-R derived from NMR spectra has a similar structural feature to part of the X2 module³⁸.

Comparison of eBCMA with eBAFF-R

The sequence homology between eBCMA and eTACI (the extracellular domain of TACI) is obvious, in contrast to that between eBCMA and eBAFF-R or between eTACI and eBAFF-R. As mentioned above, eBCMA and eBAFF-R have a similar saddle-like fold. Judged by structure superposition, the A1 module from eBCMA and the X2 module from eBAFF-R (with a r.m.s.d. of 0.6 Å) are almost identical (Fig. 3a, b). Although we could not find a corresponding D2 module in eBAFF-R for comparison, the initial one-turn helix motifs in both D2 of eBCMA and the N module of eBAFF-R are identical. This is also true for the C2 module of TNF-R1 (Fig. 3c). The major difference between eBCMA and eBAFF-R occurs after the one-turn helix. The lone extended coil in eBAFF-R contrasts with other typical TNF receptor structural modules (Fig. 3a).

The structures of eBCMA and eBAFF-R allowed us to perform a structure-based sequence alignment of eBCMA, eBAFF-R, eTACI and C2 of TNF-R1. A strong pattern of similarity emerged from this comparison (Fig. 3d). The high sequence similarities between the two CRDs in TACI and CRDs in BCMA suggest that each CRD of TACI contains one A1 module and one D2 module. Another TNF receptor member, Fn14, which also contains one CRD, was predicted to contain one A1 module and one C2 module³⁷. From our sequence-alignment result, Fn14 could contain either D2 or C2 (Fig. 3d). The ambiguity will only be resolved by a structural determination of this module. Nevertheless, we speculate that the

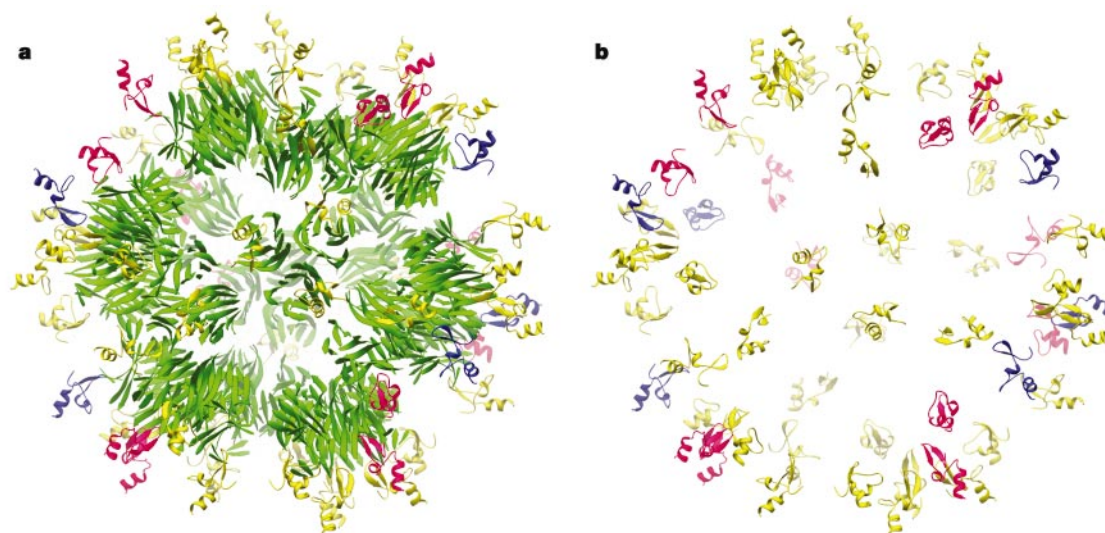


Figure 2 Overall structure of cBCMA with or without sTALL-1. **a**, The 60 monomers of sTALL-1 (green) and 60 monomers of eBCMA (molecules coloured yellow exist in the

structure, molecules coloured blue are partly ordered, and molecules coloured red are missing from the complex owing to crystal packing). **b**, As **a**, but without sTALL-1.

interaction mode between Fn14 and its ligand TWEAK are similar to what we found in the complexes of eBCMA and eBAFF-R with sTALL-1 (see below).

Interactions of sTALL-1 and eBCMA

The interactions between sTALL-1 and eBCMA are mostly in a one-to-one mode—one monomer of the receptor to one monomer of the ligand. The slightly tilted saddle-like receptor is sitting on a horse-back-like surface. This surface is formed by four loops from the ligand (the two connecting strands GH and A' A on one side and the other two connecting strands CD and EF on the other; Fig. 4a, b). This mode of interaction is markedly different from that seen within other TNF family members, in which one elongated receptor binds

to the cleft formed by two ligands (Supplementary Information). The interactions of eBCMA with sTALL-1 include hydrogen bonds, salt bridges and, most importantly, hydrophobic contacts with a contact distance of between 3 and 4 Å. A total of 21 residues are involved: 9 from eBCMA (Tyr 13, Asp 15, Leu 17, Leu 18, His 19, Ile 22, Leu 26, Arg 27 and Pro 34), 8 from the primary ligand (Tyr 163, Tyr 206, Leu 211, Arg 231, Ile 233, Pro 264, Arg 265 and Glu 266) and 4 from the secondary ligand (Leu 200, Leu 240, Asp 273 and Asp 275) of the trimer (Fig. 4c–f). The overall interactions can be divided into four groups. First, Leu 17 and Leu 18 from eBCMA, together with Tyr 163, Leu 211, Ile 233 and Pro 264 from the primary ligand and Leu 200 from the secondary ligand, form the first hydrophobic core (Fig. 4d). Tyr 163, Leu 200, Ile 233 and Pro 264 are located on a curved track. Together with Leu 17 from eBCMA, these residues form a perfect hydrophobic curved track. Leu 211 from the primary ligand and Leu 18 from eBCMA strengthen the contacts further (Fig. 4d). Second, Ile 22 and Leu 26 from eBCMA, together with Tyr 206 from the primary ligand and Leu 240 from the secondary ligand, form the second hydrophobic core (Fig. 4e). Third, Asp 15 from eBCMA and Arg 265 from the primary ligand form one salt bridge, and Arg 27 from eBCMA and Glu 266 from the primary ligand form another (Fig. 4f). There is also a potential hydrogen bond between Tyr 206 from the primary ligand and Tyr 13 from eBCMA (Fig. 4f). Last, His 19 from eBCMA forms one water-molecule-mediated interaction with Arg 231 of the primary ligand (Fig. 4d). Twenty per cent of the sTALL-1 exposed surface (3,696 Å² of the 17,868-Å² sTALL-1 trimer surface) and 38% of the eBCMA surface (1,232 Å² of the 3,237-Å² eBCMA monomer surface) are involved in the interaction.

Interactions of sTALL-1 and eBAFF-R

Most interactions between sTALL-1 and eBAFF-R are similar to those between sTALL-1 and eBCMA, although the details vary. The interactions also include hydrogen bonds, salt bridges and hydrophobic contacts. There are 24 residues involved in total: 9 from eBAFF-R (Asp 26, Leu 28, Val 29, Arg 30, Val 33, Leu 37, Leu 38, Arg 39 and Arg 42), 10 from the primary ligand (Tyr 163, Asp 203, Tyr 206, Leu 211, Arg 231, Ile 233, Pro 264, Arg 265, Glu 266 and Asn 267), 4 from the secondary ligand (Leu 200, Leu 240, Asp 273 and Asp 275), and 1 from a flap region of the neighbouring trimer (Asp 222) (Fig. 4g–i). The overall interactions also can be divided into four groups. First, Leu 28 and Val 29 from eBAFF-R, together with Tyr 163, Leu 211, Ile 233 and Pro 264 from the primary ligand and Leu 200 from the secondary ligand, form the first hydrophobic core (Fig. 4g). In comparison with eBCMA, Val 29 in eBAFF-R replaces the equivalent Leu 18. The side chain is shortened, and this could reduce the strength of the contact. Second, Val 33, Leu 37 and Leu 38 from eBAFF-R, together with Tyr 206 from the primary ligand and Leu 240 from the secondary ligand, form the second hydrophobic core (Fig. 4h). Val 33 and Leu 37 in eBAFF-R are equivalent to Ile 22 and Leu 26 in eBCMA. Leu 38 is an additional contacting residue for eBAFF-R compared with eBCMA. Third, Asp 26 from eBAFF-R and Arg 265 from the primary ligand form a salt bridge (not shown). The replacement of Arg 27 in eBCMA with Leu 38 in eBAFF-R eliminates a salt bridge with Glu 266 from the primary ligand. However, Leu 38 joins the second hydrophobic core, which might strengthen the interaction between eBAFF-R and sTALL-1 (Fig. 4h). Last, Arg 30 from eBAFF-R, Arg 231 from the primary ligand, Asp 273 and Asp 275 from the secondary ligand and Asp 222 from the third ligand form a complicated salt bridge network (Fig. 4g). The long side chain of Arg 30 from eBAFF-R (His 19 in eBCMA; Fig. 4d) makes contacts with either Asp 275 or Asp 222. The well-defined electron density of the Arg 30 side chain from eBAFF-R in the initial difference maps suggests that these are strong interactions that might considerably strengthen the eBAFF-R and sTALL-1 binding. In addition, there are three interactions over the extended coil region of eBAFF-R, which might also have a

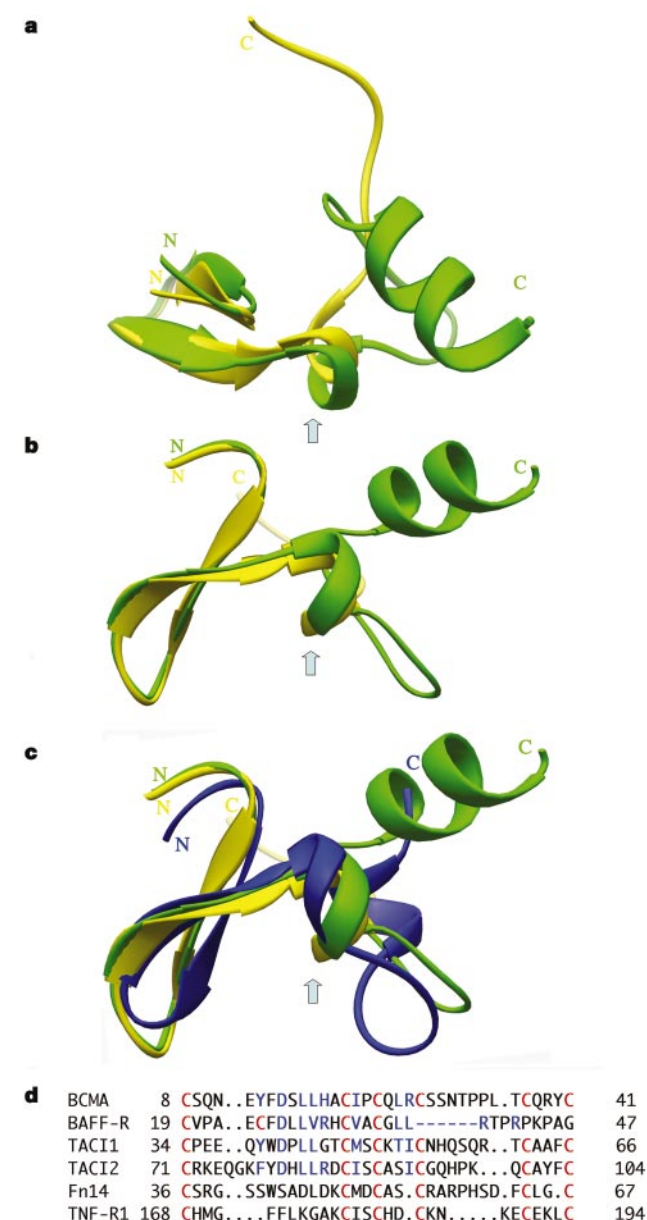


Figure 3 Comparison of different CRDs. **a**, Superposition of eBCMA (green) and eBAFF-R (yellow). **b**, Rotation of **a** through 90°. **c**, Superposition of eBCMA, eBAFF-R and the C2-containing CRD from TNF-R1 (ref. 36). **d**, Structure-based sequence alignment of CRD modules of BCMA, BAFF-R, TACI, Fn14 and TNF-R1. Residues coloured red are conserved disulphide bridges, which form modules A1, X2, D2 and C2. There is no structural similarity to other members after Leu 38 for eBAFF-R. Residues coloured blue are involved in ligand binding. Arrows indicate the one-turn helix in the N module.

crucial role in keeping the coil region in an ordered form in the structure. First, Arg 39 makes a salt bridge with Asp 203 of the primary ligand. Second, Arg 42 from eBAFF-R forms a hydrogen bond with Asn 267 from the primary ligand. Last, Glu 266 from the primary ligand makes one hydrogen bond with the main chain of the coil (Fig. 4i). In contrast, Glu 266 and Arg 27 of eBCMA make a salt bridge in the complex of eBCMA with sTALL-1 (Fig. 4f). Overall, 23% of the exposed surface of sTALL-1 ($4,197 \text{ \AA}^2$ of the $17,868 \text{ \AA}^2$ sTALL-1 trimer surface) and 44.8% of the eBAFF-R surface ($1,399 \text{ \AA}^2$ of the $3,121 \text{ \AA}^2$ eBAFF-R monomer surface) are involved in the interaction. Furthermore, eBAFF-R occupies more surface than eBCMA does on sTALL-1.

Discrimination of BAFF-R between APRIL and TALL-1

The two publications that initially reported the cloning of the BAFF-R/BR3 receptor found that BAFF-R specifically binds to TALL-1 but not to APRIL/TALL-2 (refs 14, 15). Furthermore, APRIL has a very low abundance in all tissues, and was proposed to be dispensable for B-cell maturation^{20,21}. It has also been predicted that there might be an additional and more specific

receptor for APRIL³⁹. The binding affinities of APRIL for eBCMA and eTACI are similar to those of sTALL-1 for eBCMA and eTACI³⁵. From the above structural analysis, eBCMA and eBAFF-R have nearly identical core structures that make contacts with sTALL-1. Furthermore, the interactions between eBCMA and sTALL-1 are also highly conserved in the interactions between eBAFF-R and sTALL-1. We speculate that these interactions are also conserved for TACI and sTALL-1. Given these structural similarities, why should BAFF-R discriminate between TALL-1 and APRIL? To address this question, we modelled APRIL on the basis of the sTALL-1 structure (see Methods and Supplementary Information), benefiting from the high primary-sequence homology between sTALL-1 and APRIL². The final coordinates of APRIL were superimposed on the sTALL-1 structure (Fig. 5a). Detailed interactions between eBAFF-R and APRIL were analysed. To our surprise, the interactions are extremely similar to those found in the complexes between eBCMA or eBAFF-R and sTALL-1. The first hydrophobic core we described in the two previous complexes still exists, including Leu 28 and Val 29 from eBAFF-R, Val 133, Thr 177, Val 181, Ile 197 and Pro 230 from the primary APRIL molecule, and Leu 170 from the secondary

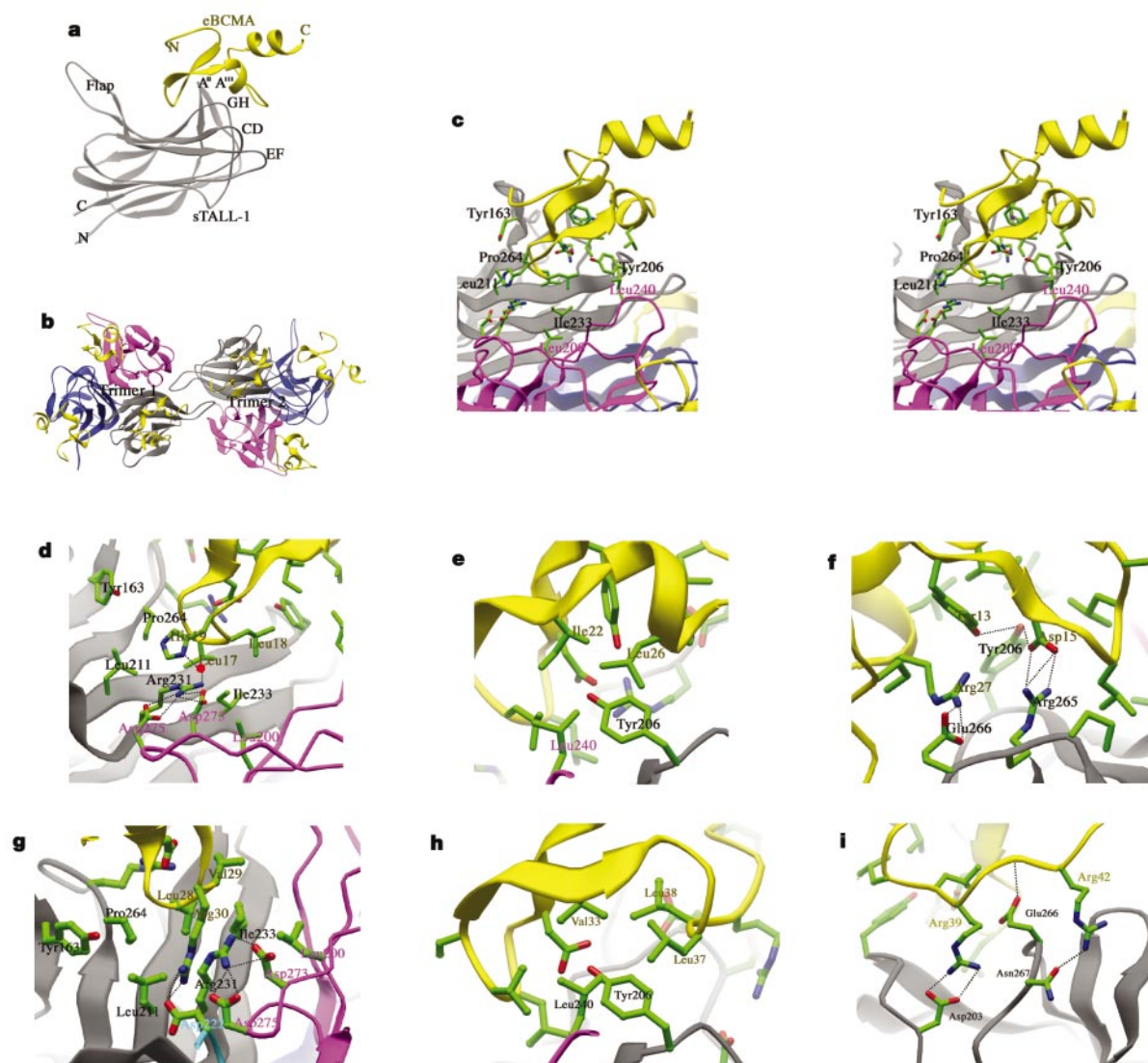


Figure 4 Detailed interactions of the two complexes. **a–f**, Interactions between eBCMA and sTALL-1: **a**, the one-to-one interaction mode of eBCMA with sTALL-1; **b**, two trimers of the complex of eBCMA and sTALL-1; **c**, the overall interactions between eBCMA and sTALL-1; **d**, hydrophobic core 1 between eBCMA and sTALL-1; **e**, hydrophobic core 2

between eBCMA and sTALL-1; **f**, salt bridges 1 and 2 in the complex of eBCMA and sTALL-1. **g–i**, Interactions between eBAFF-R and sTALL-1: **g**, hydrophobic core 1 between eBAFF-R and sTALL-1; **h**, hydrophobic core 2 between eBAFF-R and sTALL-1; **i**, hydrogen bonds and salt bridges in the extended coil region of eBAFF-R with sTALL-1.

APRIL molecule (not shown). The second hydrophobic core consists of Val 33, Leu 37, Leu 38 from eBAFF-R, Phe 176 from the primary ligand, and Arg 206 from the secondary ligand (not shown). In comparison with the interactions between eBAFF-R and sTALL-1, Tyr 206 from the primary ligand is changed to Phe 176; Leu 240 from the secondary ligand is changed to Arg 206. The former change could strengthen the hydrophobic contacts. Additionally, the major salt bridge formed by Asp 26 from the receptor and Arg 195 from APRIL is conserved (not shown). Thus, we seek the structural basis of eBAFF-R's discrimination between TALL-1 and APRIL.

To find this basis, a serial truncation, a point mutation and a combination of both were introduced (Fig. 5b). All protein versions with the above mutations were subjected to BIAcore binding assays to measure their binding affinity to sTALL-1 and APRIL. To obtain the true affinity of the receptor without the compounding effect of multivalent binding, the polyvalent sTALL-1 proteins and APRIL were immobilized in the flow cells of the instrument, and soluble monomeric eBAFF-R and its derivatives were injected in the mobile

phase. The K_d values of the interaction were calculated from the kinetic binding data. We found only one version of eBAFF-R truncation (eBAFF-R(1–36), with a complete C-terminal truncation starting from residue 37) with significant binding affinity for APRIL (Fig. 5b, c). This fragment contains only the X2 module.

These results also can be mimicked by energy minimization of models. When interaction models of APRIL and eBAFF-R are subjected to energy minimization in CNS⁴⁰, eBAFF-R is pushed away from APRIL by about 0.5 Å. This did not occur when only the X2 module was included in the process. It seems that the C-terminal half of the eBAFF-R molecule (module N) has an inhibitory role in the binding of eBAFF-R to APRIL.

The failure of eBAFF-R(1–39) to bind to APRIL (Fig. 5b) suggests that some features of residues 37–39 are important for the inhibition of binding. We therefore examined this region in eBAFF-R and eBCMA, because they are bound to sTALL-1 and we model them bound to APRIL (Fig. 5d, e). Although APRIL and sTALL-1 are structurally similar in the hydrophobic core, the microenvironment is slightly different. The aromatic side chain of Phe 176 in

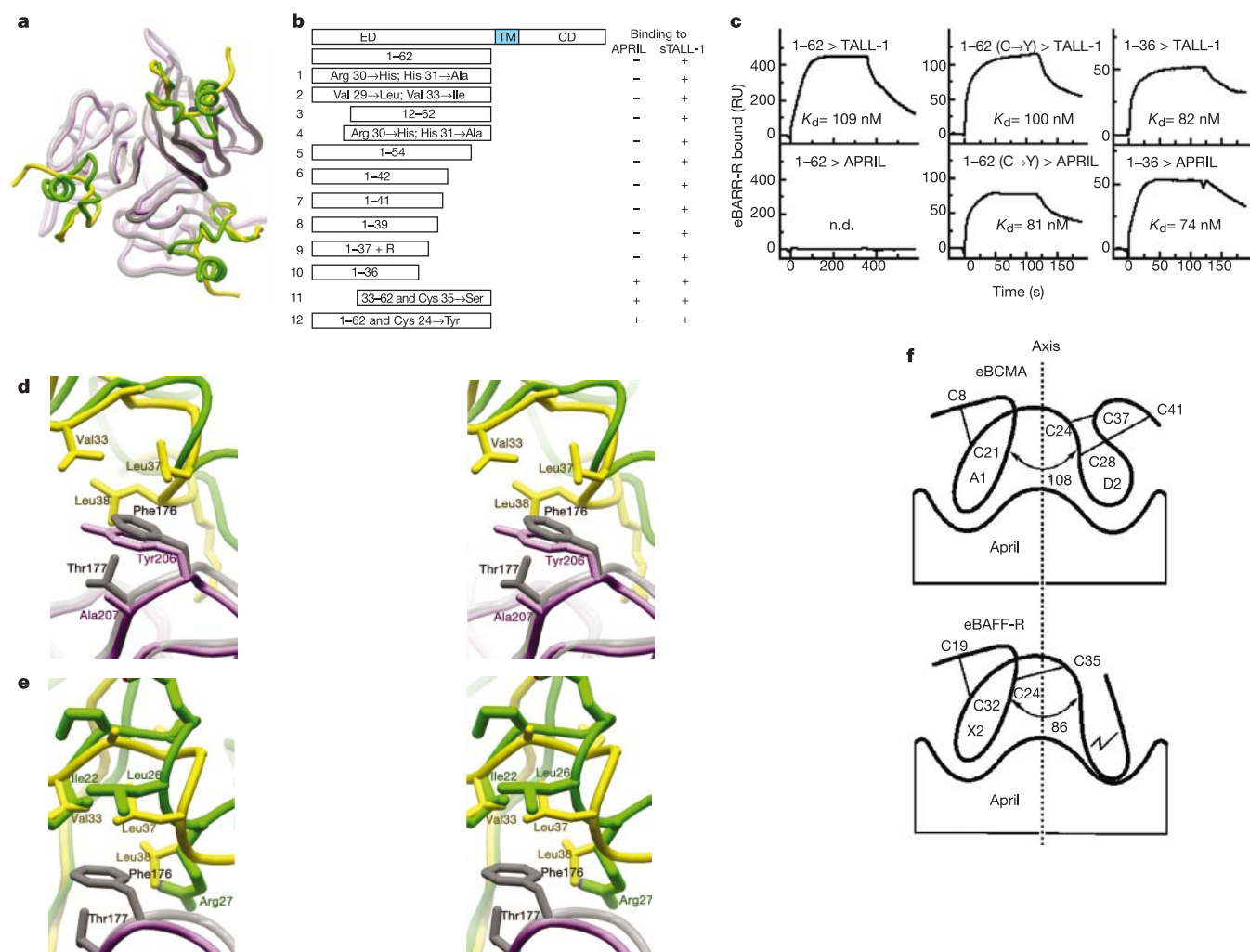


Figure 5 Characterization of ligand specificity of BAFF-R. **a**, Homology model of APRIL and its superposition on sTALL-1 in the presence of eBAFF-R (yellow) and eBCMA (green). Three sTALL-1 monomers are coloured pink; three models of APRIL are coloured light grey, grey and dark grey. **b**, Mutagenesis results of eBAFF-R. All derivatives were subjected to binding assays on APRIL with sTALL-1 as a positive control. **c**, Binding of eBAFF-R and its derivatives to sTALL-1 and APRIL. As described in Methods, surface plasmon resonance was used to assess the binding of native eBAFF-R(1–62), Cys 24 → Tyr 24 mutated eBAFF-R and truncated eBAFF-R(1–36) to immobilized sTALL-1

(upper panels) or to APRIL (lower panels). Binding data corrected for bulk refractive index are shown for 400 nM native eBAFF-R and for 1,000 nM truncated or mutant eBAFF-R. Dissociation constants (K_d) calculated from the data are also shown. **d**, Possible structural basis from the aspect of the receptor that discriminates between sTALL-1 and APRIL. Different proteins are coloured as follows: eBCMA, green; eBAFF-R, yellow; sTALL-1, pink; APRIL, dark grey. **e**, Possible structural basis from the aspect of the ligand that discriminates between sTALL-1 and APRIL, coloured as in **d**. **f**, A model of how eBAFF-R differentiates sTALL-1 from APRIL.

APRIL is predicted to be above the corresponding residue Tyr 206 in sTALL-1, which could clash with the side chain of Leu 38 in eBAFF-R (Fig. 5d). In addition, in APRIL the side chain of Thr 177 could restrict the mobility of the aromatic ring of Phe 176. This is in contrast to the more freedom given to Tyr 206 in sTALL-1 by Ala 207. From the aspect of the receptor, the one-turn helix in module N of eBAFF-R is much lower than that in eBCMA. This could lead to a repulsive contact between the one-turn helix region in eBAFF-R and Phe 176 in APRIL (Figs 3a–c and 5e).

A structural comparison between eBCMA and eBAFF-R further reveals that the unique disulphide bridge (Cys 24–Cys 35) in the X2 module creates an additional connection between modules X2 and N, which eliminates the flexibility between the two modules in eBAFF-R (Fig. 5f). On the one hand, this change pulls the two modules closer in eBAFF-R (from $\sim 108^\circ$ in eBCMA to $\sim 86^\circ$ in eBAFF-R; Fig. 5f); on the other hand, it eliminates the possibility of subtle conformation adjustment during the protein–protein recognition process because of the rigidity between two modules (Fig. 5f). To test whether the disulphide is the key component in inhibiting the binding of eBAFF-R to APRIL, we generated two point mutations, Cys 24 $\rightarrow$ Tyr 24 (the corresponding residue in eBCMA is Tyr) and Cys 35 $\rightarrow$ Ser 35. Both mutants, which cannot form the disulphide, turned out to have similar binding affinities for both sTALL-1 and APRIL (Fig. 5b, c).

Although the module N in eBAFF-R has an inhibitory role in the binding process of eBAFF-R to APRIL, we believe that module N alone should bind to sTALL-1 and APRIL, because it can freely adapt its orientation to obtain the best fit without the restraint of the disulphide bridge. A slightly modified module N of eBAFF-R (residues 33–62 and Cys 35 $\rightarrow$ Ser 35; Fig. 5b) does bind to both sTALL-1 and APRIL (data not shown). In conclusion, the above analysis reveals that the unique disulphide bridge in eBAFF-R, which restricted mobility between the X2 and N modules, is the only determinant that prevents eBAFF-R from binding to APRIL.

We have previously speculated that APRIL could have a role as a decoy ligand³¹. From the modelling results, we also found that residues involved in trimerization are absolutely conserved (data not shown). This information suggested that sTALL-1 and APRIL are expressed in the same environment and that they could simultaneously form heterotrimers. Patients with autoimmune diseases do indeed have detectable heterotrimers of sTALL-1 and APRIL in their circulation⁴¹. As we concluded previously, the sTALL-1 trimer alone (mTALL-1) is unable to trigger the signal transduction of its cognate receptors³¹. Heterotrimers of sTALL-1 and APRIL cannot form a virus-like cluster, as sTALL-1 alone does, owing to the lack of the ‘flap’ region in APRIL. We think it significant that heterotrimers of sTALL-1 and APRIL have an even higher activity than sTALL-1 alone, as reported previously⁴¹. Considering the fact that overexpression of sTALL-1 could lead to an abnormal stimulation of B cells and the development of autoimmune disease, APRIL might be serving as a balancer, reducing the opportunity for excess sTALL-1 to form the active cluster. This role is similar to the decoy death receptors, which are essential for cells to survive^{27,33,34}. Knockout data show that such mice lacking APRIL die at early embryonic stages—indirect evidence of the potentially important roles of this protein²². □

Methods

Protein expression, purification and crystallization

Protein expression, purification and crystallization of sTALL-1 have been described previously³¹. BAFF-R/BR3 was cloned as described^{14,15}. The extracellular domains of BCMA (1–54), BAFF-R (1–62) and other BAFF-R derivatives used for the experiments were overexpressed as glutathione S-transferase (GST) fusion proteins on pGEX4T-2 vector in *Escherichia coli* strain BL-21 (ref. 31). Cell preparation and protein purification procedures were similar to that of sTALL-1. In brief, harvested cells were lysed with a French press and subjected to low-speed centrifugation. The soluble fraction was loaded on GST affinity beads. After an extensive wash with binding buffer, thrombin was added and incubated for 24 h. The supernatant containing the extracellular domains of BCMA

and BAFF-R were loaded on a MonoQ column and eluted with a NaCl gradient. The protein was more than 99% pure after the MonoQ step.

Structural determination and refinement

For complex crystal preparations, sTALL-1 crystals were harvested after 2 weeks. sTALL-1 crystals were transferred to a stable soaking solution containing 40% dioxane, the corresponding receptors at 1 mM, and 100 mM Bicine, pH 9.0. After being soaked overnight, crystals were transferred to the cryoprotecting buffer system used for sTALL-1 crystals³¹. Data sets for both complexes were first collected on the in-house X-ray generator. Crystals of both complexes diffracted to 3.0 Å. A 2.6-Å data set for the complex of eBCMA with sTALL-1 and a 2.5-Å data set for the complex of eBAFF-R with sTALL-1 were collected at the Advanced Photon Source (APS) facility. All data were processed with the DENZO package⁴². Structures of the complexes were solved by difference Fourier analysis with the use of the sTALL-1 model³¹. After one round of minimization of the sTALL-1 model in CNS⁴⁰, $2F_o - F_c$ and $F_o - F_c$ maps were calculated. Models were built using O (ref. 43) and refined in CNS. CRD4 (containing modules A1 and C2) of TNF-R1 (Protein Data Bank ID 1NCF) was used for initial model building of eBCMA. Model building of eBAFF-R was helped by the available model of eBCMA (Supplementary Information).

Model building of APRIL

The most obvious difference between sTALL-1 and APRIL is in the ‘flap’ region (six residues) of sTALL-1, which APRIL lacks². We reported a mutated version of sTALL-1, with eight residues replaced by two glycine residues, that was not functional in transfection assays or in the B-cell stimulation assays but had a binding affinity to its receptors similar to that of native sTALL-1 (ref. 31). This mutated sTALL-1 is a good model for APRIL. The structure of this mutated sTALL-1 has been determined at 1.7 Å resolution by multiple isomorphous replacement; it is almost identical to that of sTALL-1, except for the missing ‘flap’ (Supplementary Information). In accordance with the sequence alignment result², the APRIL residues were mutated manually in O. The side chains of all APRIL residues were orientated as for the corresponding ones in sTALL-1. Two regions with one or two residue insertions were built automatically with O (main chain auto-building mode). The final model of APRIL was imported into the minimization program in CNS.

Protein binding assays

Wild-type multimeric sTALL-1 and APRIL (R&D Systems) were immobilized ($\sim 5,000$ resonance units (RU) of each) in separate flow cells of a CM-5 Biacore biosensor chip by using standard amine-coupling reagents. Various concentrations (100–2,000 nM) of monomeric eBAFF-R or its derivatives in a buffer containing 250 mM NaCl, 50 mM Tris buffer, pH 8.0, 20 mM EDTA and 0.005% P20 detergent were injected for 2–5 min at a flow rate of $25 \mu\text{L min}^{-1}$, and the binding kinetics were recorded. To correct for bulk refractive index, eBAFF-R and its derivative samples were injected into a control flow cell with no protein immobilized. Association and dissociation rates were calculated with standard Biacore software (BIAevaluation 3.0). For both sTALL-1 and APRIL experiments, the model $A + B = AB$ and the fitting option correcting for mass transport effects were used.

Received 27 January; accepted 11 March 2003; doi:10.1038/nature01543.

- Locksley, R. M., Killeen, N. & Lenardo, M. J. The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* **104**, 487–501 (2001).
- Shu, H. B., Hu, W. H. & Johnson, H. TALL-1 is a novel member of the TNF family that is downregulated by mitogens. *J. Leukocyte Biol.* **65**, 680–683 (1999).
- Schneider, P. et al. BAFF, a novel ligand of the tumour necrosis factor family, stimulates B cell growth. *J. Exp. Med.* **189**, 1747–1756 (1999).
- Moore, P. A. et al. BlyS: member of the tumour necrosis factor family and B lymphocyte stimulator. *Science* **285**, 260–263 (1999).
- Mukhopadhyay, A., Ni, J., Zhai, Y., Yu, G. L. & Aggarwal, B. B. Identification and characterization of a novel cytokine, THANK, a TNF homologue that activates apoptosis, nuclear factor- κ B, and c-Jun NH₂-terminal kinase. *J. Biol. Chem.* **274**, 15978–15981 (1999).
- Shu, H. B. & Johnson, H. B cell maturation protein is a receptor for the TNF family member TALL-1. *Proc. Natl Acad. Sci. USA* **97**, 9156–9161 (2000).
- Gross, J. A. et al. TACI and BCMA are receptors for a TNF homologue implicated in B cell autoimmune disease. *Nature* **404**, 995–999 (2000).
- Thompson, J. S. et al. BAFF binds to the tumour necrosis factor receptor-like molecule B cell maturation antigen and is important for maintaining the peripheral B cell population. *J. Exp. Med.* **192**, 129–135 (2000).
- Marsters, S. A. et al. Interaction of the TNF homologues BlyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* **10**, 785–788 (2000).
- Xia, X. Z. et al. TACI is a TRAF-interacting receptor for TALL-1, a tumour necrosis factor family member involved in B cell regulation. *J. Exp. Med.* **192**, 137–143 (2000).
- Yan, M. et al. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* **1**, 37–41 (2000).
- Mackay, F. et al. Mice transgenic for BAFF develop lymphocytic disorders along with autoimmune manifestations. *J. Exp. Med.* **190**, 1697–1710 (1999).
- Khare, S. D. et al. Severe B cell hyperplasia and autoimmune disease in TALL-1 transgenic mice. *Proc. Natl Acad. Sci. USA* **97**, 3370–3375 (2000).
- Thompson, J. S. et al. BAFF-R, a novel TNF receptor that specifically interacts with BAFF. *Science* **293**, 2108–2111 (2001).
- Yan, M. et al. Identification of a novel receptor for B lymphocyte stimulator that is mutated in a mouse strain with severe B cell deficiency. *Curr. Biol.* **11**, 1547–1552 (2001).
- Schiemann, B. et al. An essential role for BAFF in the normal development of B cells through a BCMA-independent pathway. *Science* **293**, 2111–2114 (2001).
- Gross, J. A. et al. TACI-Ig neutralizes molecules critical for B-cell development and autoimmune disease. Impaired B-cell maturation in mice lacking BlyS. *Immunity* **15**, 289–302 (2001).

18. Xu, S. & Lam, K.-P. B-cell maturation protein, which binds the tumour necrosis factor family members BAFF and APRIL, is dispensable for humoral immune response. *Mol. Cell. Biol.* **21**, 4067–4074 (2001).
19. Von Bulow, G. U., van Deursen, J. M. & Bram, R. J. Regulation of the T-independent humoral response by TACI. *Immunity* **14**, 573–582 (2001).
20. Hahne, M. *et al.* APRIL, a new ligand of the tumour necrosis factor family, stimulates tumour cell growth. *J. Exp. Med.* **188**, 1185–1190 (1998).
21. Schneider, P. *et al.* Maturation of marginal zone and follicular B cells requires B cell activation factor of the tumour necrosis factor family and is independent of B cell maturation antigen. *J. Exp. Med.* **194**, 1691–1697 (2001).
22. Mackay, F. & Mackay, C. R. The role of BAFF in B-cell maturation, T-cell activation and autoimmunity. *Trends Immunol.* **23**, 113–115 (2002).
23. Jones, E. Y., Stuart, D. I. & Walker, N. P. Structure of tumour necrosis factor. *Nature* **338**, 225–228 (1989).
24. Eck, M. J. & Sprang, S. R. The structure of tumour necrosis factor- α at 2.6 Å resolution: implications for receptor binding. *J. Biol. Chem.* **264**, 17595–17605 (1989).
25. Eck, M. J., Ultsch, M., Rinderknecht, E., de Vos, A. M. & Sprang, S. R. The structure of human lymphotoxin (tumour necrosis factor β) at 1.9 Å resolution. *J. Biol. Chem.* **267**, 2119–2122 (1992).
26. Karpusas, M. *et al.* 2 Å crystal structure of an extracellular fragment of human CD40 ligand. *Structure* **3**, 1031–1039 (1995).
27. Cha, S. S. *et al.* 2.8 Å resolution crystal structure of human TRAIL, a cytokine with selective antitumor activity. *Immunity* **11**, 253–261 (1999).
28. Lam, J., Nelson, C. A., Ross, F. P., Teitelbaum, S. L. & Fremont, D. H. Crystal structure of the TRANCE/RANKL cytokine reveals determinants of receptor–ligand specificity. *J. Clin. Invest.* **108**, 971–979 (2001).
29. Karpusas, M. *et al.* Crystal structure of extracellular human BAFF, a TNF family member that stimulates B lymphocytes. *J. Mol. Biol.* **315**, 1145–1154 (2002).
30. Oren, D. A. *et al.* Structural basis of BlyS receptor recognition. *Nature Struct. Biol.* **9**, 288–292 (2002).
31. Liu, Y. *et al.* Crystal structure of sTALL-1 reveals a virus-like assembly of TNF family ligands. *Cell* **108**, 383–394 (2002).
32. Banner, D. W. *et al.* Crystal structure of the soluble human 55 kd TNF receptor–human TNF complex: implications for TNF receptor activation. *Cell* **73**, 431–445 (1993).
33. Mongkolsapaya, J. *et al.* Structure of the TRAIL–DR5 complex reveals mechanisms conferring specificity in apoptotic initiation. *Nature Struct. Biol.* **6**, 1048–1053 (1999).
34. Hymowitz, S. G. *et al.* Triggering cell death: the crystal structure of Apo2L/TRAIL in a complex with death receptor 5. *Mol. Cell* **4**, 563–571 (1999).
35. Yu, G. *et al.* APRIL and TALL-1 and receptors BCMA and TACI: system for regulating humoral immunity. *Nature Immunol.* **1**, 252–256 (2000).
36. Naismith, J. H. & Sprang, S. R. Modularity in the TNF-receptor family. *Trends Biochem. Sci.* **23**, 74–79 (1998).
37. Bodmer, J. L., Schneider, P. & Tschopp, J. The molecular architecture of the TNF superfamily. *Trends Biochem. Sci.* **27**, 19–26 (2002).
38. Kayagaki, N. *et al.* BAFF/BlyS receptor 3 binds the B cell survival factor BAFF ligand through a discrete surface loop and promotes processing of NF- κ B2. *Immunity* **17**, 515–524 (2002).
39. Ware, C. F. APRIL and BAFF connect autoimmunity and cancer. *J. Exp. Med.* **192**, F35–F37 (2000).
40. Brunger, A. T. *et al.* Crystallography and NMR System (CNS): A new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
41. Roschke, V. *et al.* BlyS and APRIL form biologically active heterotrimers that are expressed in patients with systemic immune-based rheumatic diseases. *J. Immunol.* **169**, 4314–4321 (2002).
42. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
43. Jones, T. A., Zou, J.-Y., Cowan, S. & Kjeldgaard, A. L. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
44. Carson, M. Ribbon models of macromolecules. *Methods Enzymol.* **277**, 493–505 (1997).
45. Esnouf, R. M. BobScript v2.4 changes (C). *J. Mol. Graphics* **15**, 132–134 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Zhao, O. Swartz and L. Halton for reading our manuscript; Z. Chen for Fig. 5f; M. Carson, S. R. Sprang, C. R. Mackay and J. Tschopp for their suggestions; the Howard Hughes Medical Institute, Zuckerman/Canyon Ranch and A. Lapporte for the support of our X-ray and computing facility; A. Joachimiak and SBC (ID-19) at APS for high-resolution data; and P. C. Marrack, J. D. Crapo, J. Cambier, X. Liu and other members at the National Jewish Medical and Research Center for support. S.H.B. is supported by an NIH grant and the Arthritis Foundation; Y.L. is supported partly by a Priscilla Campbell Memorial Fellowship; X.H. is supported partly by a Janet S. Lewald fellowship; G.Z. is supported by a PEW Scholar Award, a start-up fund from National Jewish Medical and Research Center and the Cancer League of Colorado, Inc.; P.M. and G.Z. are supported by NIH grants.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.Z. (zhangg@njc.org). The atomic coordinates and reflections have been deposited in the Protein Data Bank under accession codes 1OQD and 1OQE.

The elemental abundance pattern in a galaxy at $z = 2.626$

Jason X. Prochaska*, J. Christopher Howk† & Arthur M. Wolfe‡

* UCO/Lick Observatory, University of California, Santa Cruz, California 95064, USA

† Department of Physics, and Center for Astrophysics and Space Sciences, University of California, San Diego, C-0424, La Jolla, California 92093, USA

The discovery of metal-poor stars^{1,2} (where metal is any element more massive than helium) has enabled astronomers to probe the chemical enrichment history of the Milky Way^{3,4}. More recently, element abundances in gas inside high-redshift galaxies has been probed through the absorption lines imprinted on the spectra of background quasars^{5–8}, but these have typically yielded measurements of only a few elements. Furthermore, interpretation of these abundances is complicated by the fact that differential incorporation of metals into dust can produce an abundance pattern similar to that expected from nucleosynthesis by massive stars⁹. Here we report the observation of over 25 elements in a galaxy at redshift $z = 2.626$. With these data, we can examine nucleosynthetic processes independent of the uncertainty arising from depletion. We find that the galaxy was enriched mainly by massive stars ($M > 15$ solar masses) and propose that it is the progenitor of a massive elliptical galaxy. The detailed abundance patterns suggest that boron is produced through processes that act independently of metallicity, and may require alternative mechanisms for the nucleosynthesis of germanium.

With the exception of the light elements (Li, Be, B and partly He) whose origin is linked to either Big Bang nucleosynthesis¹⁰ or unique astrophysical processes¹¹, the elements up to Fe on the periodic table are produced during the nuclear reactions that fuel stars or during the collapse of the stellar core and the resulting explosion defining a supernova¹². Regarding supernovae, the current physical picture states that even-atomic-number (even- Z) nuclei—including the ‘ α -elements’ O, Ne, Mg, Si, S, Ar, and Ca—are synthesized by the supernovae of massive stars, whereas a substantial component of the Fe-peak elements are produced on significantly longer timescales in supernovae with lower mass progenitors. Therefore, comparisons between the α -elements and the Fe-peak clarify the star-formation history and age of the galaxy¹³. The heavier elements—whose nuclei are energetically less favourable than Fe—require a significant source of free neutrons and the sequence of elements and isotopes created is very sensitive to the neutron capture rate relative to β -decay: the s -process (r -process) refers to a slow (fast) rate. To our knowledge, these heavier elements have never been detected outside our Galaxy and its nearest neighbours.

The galaxy discussed here was discovered¹⁴ in absorption at $z = 2.626$ along the sightline to the background quasar FJ081240.6+320808 (emission redshift $z_{\text{em}} = 2.701$) via the signature damping wings of the Ly α profile that give the galaxy its classification—a damped Ly α system¹⁵ (DLA). We obtained a moderate-resolution spectrum of FJ081240.6+320808 with the Echelle Spectrograph and Imager¹⁶ on the Keck II telescope as part of a larger programme to survey the chemical enrichment history of the Universe¹⁷. The unusually strong metal-line absorption profiles of this galaxy led us to acquire follow-up observations of FJ081240.6+320808 with the HIRES echelle spectrograph¹⁸ on Keck I. Our observations revealed a series of metal-line transitions (Fig. 1) previously undetected at high redshift which provide the most comprehensive set of elemental abundances available. By performing standard line-profile fitting and equivalent width measurements, we have measured gas-phase abundances for the

elements listed in Table 1. (See also the database at <http://kingpin.ucsd.edu/~hiresdla/>.) This DLA has the highest combined H I column density and metallicity ($O/H \approx 1/3$ solar abundance) ever observed. We note that its redshift implies a strict upper limit to its age of 2.5 Gyr in any realistic cosmology.

The gas-phase abundances listed in Table 1 qualitatively follow the differential depletion pattern observed along the sightlines through the interstellar medium of the Milky Way galaxy¹⁹. Refractory elements like Fe, Ni and Cr must be highly depleted onto dust grains in this high redshift galaxy. The presence of strong C II* and Cl I absorption suggests that the gas resides in a cold neutral medium characteristic of highly depleted gas in the Milky Way. Furthermore, the observation of significant Cl I requires at least a modest molecular hydrogen fraction²⁰. To examine the underlying enrichment history of this galaxy we have applied empirical, conservative dust corrections (Table 1) derived from the depletion patterns of the Milky Way to the gas-phase abundances. Although these corrections are based on our limited knowledge of dust in the local Universe, we stress that the following discussion of nucleosynthesis is not sensitive to the corrections unless otherwise noted.

Figure 2 reveals the nucleosynthetic pattern of this galaxy where the dotted line compares the solar abundance pattern scaled to the oxygen metallicity of the galaxy. At the crudest level, the galaxy’s enrichment pattern resembles the Sun, indicating that their nucleosynthetic enrichment histories are not too dissimilar. Quantitatively, one might expect that the galaxy’s young age and relatively high metallicity would imply a nucleosynthetic pattern dominated by short-lived, massive stars. This presumption is supported by several lines of evidence. First, the α -elements (O, Mg, Si) exhibit enhanced relative abundances compared to Zn, the only element representative of the Fe-peak not strongly incorporated into dust grains; O, Mg and Si are therefore presumably enhanced relative to the entire Fe-peak. This α /Zn enhancement is suggestive of enrichment by massive stars. Second, the relative abundances of the α -elements exhibit a trend of lower relative abundance at higher Z (for example, $[O/S] \approx -0.3$). This trend reflects the detailed pro-

Table 1 Chemical abundances for the galaxy at $z = 2.626$

Element	$[X/H]^*$	$\delta_{\text{dust}}^\dagger$	$[X/O]^\ddagger$
B	−0.57	0.1	−0.03
N	>−2.24	0.0	>−1.80
O	−0.54	0.1	0.00
Mg	−0.78	0.3	−0.04
Al	>−2.00	>0.5	>−1.06
Si	−0.91	0.3	−0.17
P	<−1.06	<0.3	<−0.32
S	−0.87	0.1	−0.33
Cl	−1.55	>0.0	>−1.11
Ti	−1.87	>0.7	>−0.73
Cr	−1.61	>0.7	>−0.47
Mn	<−1.85	0.7	<−0.71
Fe	−1.69	>0.7	>−0.55
Co	<−1.48	>0.7	>−0.34
Ni	−1.73	>0.7	>−0.59
Cu	<−1.11	>0.7	>−0.03
Zn	−0.91	0.2	−0.27
Ga	<−1.45	0.7	<−0.31
Ge	−0.92	0.3	−0.18
As	<0.26	0.0	<0.70
Kr	<−0.44	0.0	<0.00
Sr	<−0.27	0.0	<0.17
Pb	<−0.10	0.0	<0.34

* Gas-phase abundances relative to solar on a logarithmic scale; for example, $[X/H] = -1$ implies 1/10 solar abundance. Throughout the analysis we have adopted $N(\text{H I}) = 10^{21.35} \text{ cm}^{-2}$ measured from a fit to the damped Ly α profile.

† Dust corrections estimated from the depletion patterns of Galactic gas with comparable depletion levels¹⁹. The values are added to the gas-phase abundances to yield the underlying nucleosynthetic pattern. In all cases, we have considered very conservative corrections.

‡ Dust-corrected abundances on a solar logarithmic scale relative to O. These values express the nucleosynthetic pattern of this protogalaxy.

duction factors predicted for very massive stars²¹ and their supernovae. Finally, all of the odd- Z elements show subsolar relative abundances marking the enhanced ‘odd–even effect’ indicative of massive supernovae. This includes $[\text{Mn}/\text{Fe}] < -0.16$, $[\text{Ga}/\text{Ge}] < -0.13$ and $[\text{P}/\text{Si}] < -0.15$.

The high metallicity and relative abundances of this galactic gas resemble stellar abundances observed in massive, early-type galaxies (for example, ellipticals) dominated by old stellar populations²². One might speculate, then, that this galaxy is the progenitor of one of these massive structures—which could explain its high metallicity at such an early epoch. In fact, its metallicity and dust content emission at $z > 3$ that are the presumed progenitors of massive galaxies. The similarity in redshift, metallicity and relative abundances with the well-studied Lyman-break galaxy CB58 is

particularly noteworthy²⁴. Furthermore, early-type galaxies may exhibit²⁵ an overabundance of Mg/Ca , which resembles the α -element versus Z trend observed for this protogalaxy. Follow-up imaging and spectroscopy of the galaxies in the field surrounding FJ081240.6+320808 may help to reveal its stellar properties.

Our data place constraints on nucleosynthetic processes in the early Universe. The observations, for example, help to distinguish between processes where B production scales with the galaxy’s metallicity and processes that are independent of metallicity. The former theories (such as neutrino spallation in the carbon shells of supernovae²⁶ and the spallation of C and O nuclei accelerated by supernovae onto local interstellar gas²⁷) predict that the B/O ratio will remain constant with O/H metallicity while the latter theories (such as p,n accelerated onto interstellar CNO seed nuclei) predict

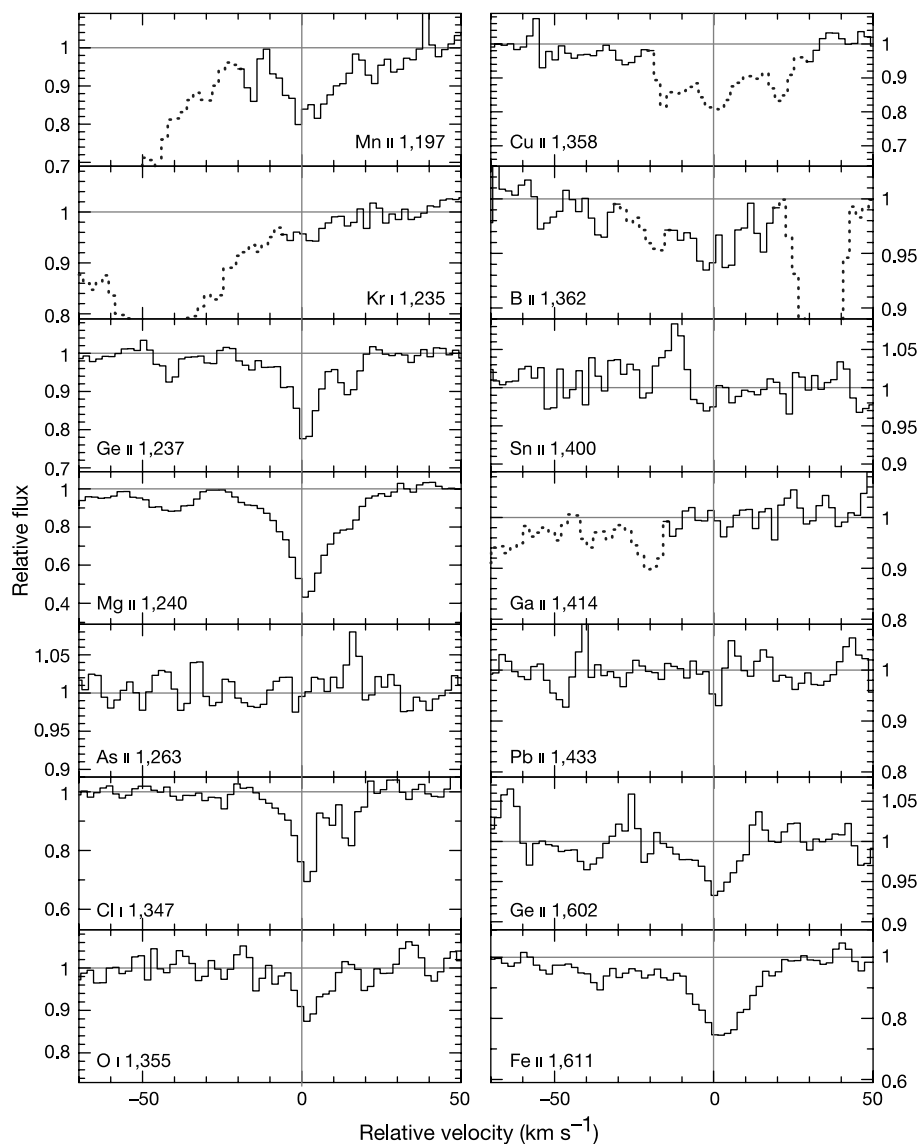


Figure 1 Sample of previously undetected metal-line transitions. Normalized absorption-line profiles related to the galaxy at $z = 2.626$ toward FJ081240.6+320808 taken with the HIRES echelle spectrograph on the Keck I 10-m telescope ($R \approx 45,000$, signal-to-noise ratio ≈ 30 per 2 km s^{-1} pixel). With the exception of $\text{Fe II } 1,611$ and $\text{Mg II } 1,240$, none of these transitions have been detected outside of our Galaxy. Line blends with coincident absorption features are designated by dotted lines. The velocity $v = 0 \text{ km s}^{-1}$ was arbitrarily defined to correspond to $z = 2.6263$. With the exception of Cl I , these

transitions correspond to the dominant ion of these elements in neutral hydrogen gas. Therefore, we convert the ionic column densities measured from these transitions into elemental abundances without ionization corrections. Although several transitions provide only upper limits to the elemental abundance (for example, $\text{Pb II } 1,433$, $\text{Ga II } 1,414$), the spectral regions are free from line-blends and future observations will yield detections or important limits.

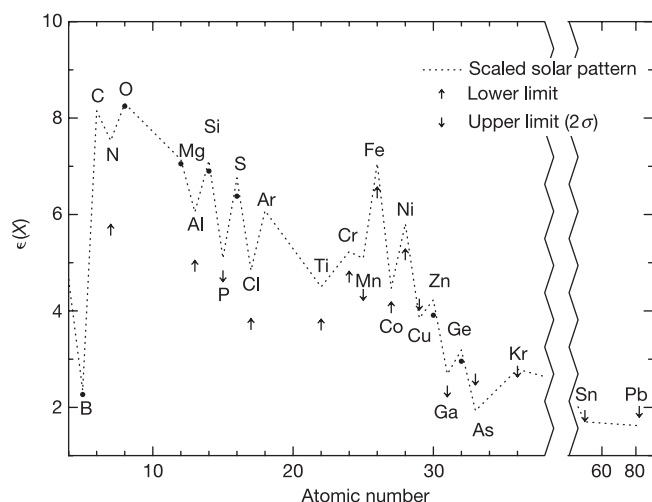


Figure 2 The nucleosynthetic enrichment pattern for a galaxy discovered in the early Universe. Dust-corrected elemental abundances for the protogalaxy at $z = 2.626$ toward FJ081240.6+320808 on a logarithmic scale where hydrogen is defined to have abundance number $\epsilon(X) = 12$. The dust corrections were empirically derived from depletion patterns observed in the local Universe and were applied in a very conservative manner. This explains the lower limits to the Fe, Cr, Al, Co and Ni abundances and similarly the upper limit to Mn. Typical statistical errors are <0.1 dex, that is, the size of the plot symbols (circles). The dotted line traces the solar abundance pattern scaled to match the observed oxygen abundance ($[O/H] = -0.44$, after dust correction). To zeroth order, the pattern of this high-redshift galaxy resembles that of the Sun; at finer detail, we note several important differences: see text for discussion.

that the B/O ratio will increase with increasing metallicity. The current observation of a solar B/O ratio in this approximately 1/3 solar metallicity gas argues for B synthesis independent of metallicity. Future observations could test this hypothesis in other galaxies and may allow a measurement of the $^{10}\text{B}/^{11}\text{B}$ isotopic ratio, which will be critical to determining the relative importance of neutrino and cosmic ray spallation¹².

In addition, the nearly solar Ge/O ratio we observe raises new challenges for the production of Ge. Although the s-process is expected to be a principal channel for Ge nucleosynthesis, the primary s-process site is during the 'asymptotic giant branch' phase of low-mass stars²⁸ whose lifetimes may exceed the age of this young galaxy. In addition to the s-process, supernovae with low-mass progenitors may produce a significant fraction of the Ge observed in our Galaxy. We have argued, however, that this high redshift galaxy is dominated by enrichment from massive stars, so the observed Ge/O ratio may support distinct physical processes (such as the neutrino-wind process²⁹). These and related issues may be investigated by detecting or significantly lowering the current limits for Pb, Sn and Ga; Pb and Sn production are dominated by the s-process, whereas Ga may be produced in a similar fashion to Ge.

This galaxy also exhibits transitions of C I , C I^* , C I^{**} , C II , C II^* , and Si II^* that will yield an unparalleled analysis of its physical properties³⁰ (for example, ρ , T , pressure, star formation rate) and an assessment of the cosmic microwave background temperature at $z = 2.6$. These observations will enable us to characterize the physical characteristics and star formation history of this young galaxy as well as provide detections or important limits on Ar, Ti, C, N, P, Pb, Sn, and possibly Kr and Fe. We could then test theories related to CNO and r-process nucleosynthesis.

Absorption-line studies of DLA provide the only means of examining elements beyond the Fe-peak outside the very local

Universe. The discovery of this galaxy indicates that similar abundance analyses will be possible in a small but non-negligible sample (about 2%) of high-redshift DLA. Finally, we report the discovery of a second DLA which should present an abundance analysis of about 20 elements including Ga, Ge, Cu and possibly B. This DLA is identified along the same sightline to FJ081240.6+320808 but at a cosmologically distinct redshift. □

Received 10 December 2002; accepted 24 February 2003; doi:10.1038/nature01524.

- Chamberlain, J. W. & Aller, L. H. The atmospheres of A-type subdwarfs and 95 Leonis. *Astrophys. J.* **114**, 52–72 (1951).
- Helfer, H. L., Wallerstein, G. & Greenstein, J. L. Abundances in some population II K giants. *Astrophys. J.* **129**, 700–719 (1959).
- Wheeler, J. C., Sneden, C. & Truran, J. W. Jr Abundance ratios as a function of metallicity. *Annu. Rev. Astron. Astrophys.* **27**, 279–349 (1989).
- McWilliam, A., Preston, G. W., Sneden, C. & Searle, L. A spectroscopic analysis of 33 of the most metal-poor stars. I. *Astron. J.* **109**, 2757–2799 (1995).
- Lu, L., Sargent, W. L. W., Barlow, T. A., Churchill, C. W. & Vogt, S. Abundances at high redshifts: The chemical enrichment history of damped Ly α galaxies. *Astrophys. J. Suppl.* **107**, 475–519 (1996).
- Pettini, M., Ellison, S. L., Steidel, C. C., Shapely, A. L. & Bowden, D. V. Si and Mn abundances in damped Ly α systems with low dust content. *Astrophys. J.* **532**, 65–76 (2000).
- Molaro, P. et al. UVES observations of QSO 0000–2620: Oxygen and zinc abundances in the damped Ly α galaxy at $z = 3.3901$. *Astrophys. J.* **541**, 54–60 (2000).
- Prochaska, J. X. & Wolfe, A. M. The UCSD HIREs/Keck I damped Ly α abundance database. II. The implications. *Astrophys. J.* **566**, 68–92 (2002).
- Vladilo, G. Chemical abundances of damped Ly α systems: A new method for estimating dust depletion effects. *Astron. Astrophys.* **391**, 407–415 (2002).
- Schramm, D. N. & Turner, M. S. Big-bang nucleosynthesis. *Rev. Mod. Phys.* **70**, 303–318 (1998).
- Fields, B. D. & Olive, K. A. The revival of galactic cosmic-ray nucleosynthesis? *Astrophys. J.* **516**, 797–810 (1999).
- Burbidge, E. M., Burbidge, G. R., Fowler, W. A. & Hoyle, F. Synthesis of the elements in stars. *Rev. Mod. Phys.* **29**, 547–650 (1957).
- Tinsley, B. M. Stellar lifetimes and abundance ratios in chemical evolution. *Astrophys. J.* **229**, 1046–1056 (1979).
- White, R. L. et al. The FIRST bright quasar survey II. 60 nights and 1200 spectra later. *Astrophys. J. Suppl.* **126**, 133–207 (2000).
- Wolfe, A. M., Turnshek, D. A., Smith, H. E. & Cohen, R. D. Damped Lyman-alpha absorption by disk galaxies with large redshifts. I.—The Lick survey. *Astrophys. J. Suppl.* **61**, 249–304 (1986).
- Sheinis, A. I. et al. SI, a New Keck Observatory echelle spectrograph and imager. *Proc. Astron. Soc. Pacif.* **114**, 851–865 (2002).
- Prochaska, J. X., Gawiser, E., Wolfe, A. M., Cooke, J. & Gelino, D. The ESI/Keck II damped Ly α abundance database. *Astrophys. J. Suppl.* (submitted).
- Vogt, S. S. et al. HIREs: the high-resolution echelle spectrometer on the Keck 10-m telescope. *SPIE* **2198**, 362–375 (1994).
- Savage, B. D. & Sembach, K. R. Interstellar abundances from absorption-line observations with the Hubble Space Telescope. *Annu. Rev. Astron. Astrophys.* **34**, 279–330 (1996).
- Jura, M. & York, D. G. Observations of interstellar chlorine and phosphorus. *Astrophys. J.* **219**, 861–869 (1978).
- Woosley, S. E. & Weaver, T. A. The evolution and explosion of massive stars. II. Explosive hydrodynamics and nucleosynthesis. *Astrophys. J. Suppl.* **101**, 181–235 (1995).
- Trager, S. C., Faber, S. M., Worthey, G. & González, J. J. The stellar population histories of local early-type galaxies. I. Population parameters. *Astron. J.* **119**, 1645–1676 (2000).
- Pettini, M. et al. The rest-frame optical spectra of Lyman break galaxies: Star formation, extinction, abundances, and kinematics. *Astrophys. J.* **554**, 981–1000 (2001).
- Pettini, M. et al. The ultraviolet spectrum of MS 1512–CB58: An insight into Lyman-break galaxies. *Astrophys. J.* **528**, 96–107 (2000).
- Saglia, R. P., Maraston, C., Thomas, D. & Bender, R. The puzzlingly small Ca II triplet absorption in elliptical galaxies. *Astrophys. J.* **579**, L13–L16 (2002).
- Woosley, S. E., Hartmann, D., Hoffman, R. D. & Haxton, W. The ν -process. *Astrophys. J.* **356**, 272–301 (1990).
- Cassé, M., Lehoucq, R. & Vangioni-Flam, E. Production and evolution of light elements in active star-forming regions. *Nature* **373**, 318–321 (1995).
- Busso, M., Gallino, R. & Wasserburg, G. J. Nucleosynthesis in asymptotic giant branch stars: relevance for Galactic enrichment and Solar System formation. *Annu. Rev. Astron. Astrophys.* **37**, 239–309 (1999).
- Hoffman, R. D., Woosley, S. E., Fuller, G. M. & Meyer, B. S. Production of the light p-process nuclei in neutrino-driven winds. *Astrophys. J.* **460**, 478–488 (1996).
- Wolfe, A. M., Prochaska, J. X. & Gawiser, E. CII* absorption in damped Ly α systems: A new window on the star formation history of the universe. *Astrophys. J.* (submitted).

Acknowledgements These observations were made with the W.M. Keck Telescope. The Keck Observatory is a joint facility of the University of California, the California Institute of Technology, and NASA. J.C.H. acknowledges support from a NASA grant to UC San Diego.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.X.P. (xavier@ucolick.org).

Interplanetary dust from the explosive dispersal of hydrated asteroids by impacts

Kazushige Tomeoka*, Koji Kiriya*, Keiko Nakamura*†, Yasuhiro Yamahana* & Toshimori Sekine‡

* Department of Earth and Planetary Sciences, Faculty of Science, Kobe University, Nada, Kobe 657-8501, Japan

‡ National Institute for Materials Science, Tsukuba 305-0044, Japan

† Present address: Office of Astromaterials Research and Exploration Science, NASA JSC, Houston, Texas 77058, USA

The Earth accretes about 30,000 tons of dust particles per year, with sizes in the range of 20–400 μm (refs 1, 2). Those particles collected at the Earth's surface—termed micrometeorites—are similar in chemistry and mineralogy to hydrated, porous meteorites^{3–7}, but such meteorites comprise only 2.8% of recovered falls⁸. This large difference in relative abundances has been attributed to 'filtering' by the Earth's atmosphere⁹, that is, the porous meteorites are considered to be so friable that they do not survive the impact with the atmosphere. Here we report shock-recovery experiments on two porous meteorites, one of which is hydrated and the other is anhydrous. The application of shock to the hydrated meteorite reduces it to minute particles and explosive expansion results upon release of the pressure, through a much broader range of pressures than for the anhydrous meteorite. Our results indicate that hydrated asteroids will produce dust particles during collisions at a much higher rate than anhydrous asteroids, which explains the different relative abundances of the hydrated material in micrometeorites and meteorites: the abundances are established before contact with the Earth's atmosphere.

The carbonaceous chondrites subtypes CI and CM are the hydrated, porous meteorites that represent, chemically and miner-

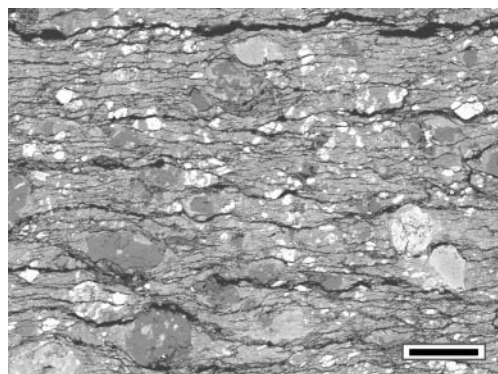


Figure 1 Back-scattered SEM image of a portion of the Murchison sample shocked at 28 GPa. Dense, narrow, subparallel fractures in the matrix are roughly perpendicular to the compression axis. In this image and those in Figs 2 and 4, the vertical direction is parallel to the compression axis. Scale bar, 100 μm . The Murchison samples were shocked in nine experiments at peak pressures 4, 10, 21, 26, 28, 30, 34, 36 and 49 GPa. The target samples are disks of meteorites that are 5.7–10.0 mm in diameter and 1.0–3.5 mm in thickness. The velocities of projectiles range from 0.50 to 1.80 km s^{-1} . The details of shock experimental procedures are described in ref. 30. The present results on Murchison are from the materials recovered from the previous experiments³⁰, whereas those on Allende are from new experiments. The samples were studied using an optical microscope, an SEM equipped with an energy dispersive X-ray spectrometer (EDS), and an electron microprobe analyser.

alogically, the most primitive material in the Solar System. They consist in large part of matrix that is composed mainly of fine-grained phyllosilicates, which account for most of the water in the meteorites (10–20 wt% H_2O).

A series of shock-recovery experiments have been performed on the Murchison CM and Allende CV carbonaceous chondrites using a single-stage 30-mm bore propellant gun. Both Murchison and Allende contain major volume fractions of matrix (~64 and ~38 vol.%, respectively^{10,11}). The matrix of Murchison consists mainly of hydrous serpentine and tochilinite, whereas the matrix of Allende consists mainly of anhydrous olivine. The bulk H_2O content for Murchison is ~12 wt% (ref. 12), whereas that for Allende is negligible. The porosities of bulk Murchison and Allende are 23% and 20%, respectively¹³. The Murchison samples were shocked in nine experiments at peak pressures from 4 to 49 GPa, whereas the Allende samples were shocked in five experiments at peak pressures from 27 to 49 GPa.

Scanning electron microscope (SEM) observations show that the matrix of Murchison shocked at pressures of less than 21 GPa exhibits sparse fractures (1–20 μm in width). At 21 GPa, minor amounts of local melts occur as veins and pockets. Otherwise, no marked changes are observed in the matrix. At 26–30 GPa, however, the matrix exhibits very noticeable changes. Narrow, subparallel fractures form at high density in directions roughly perpendicular to the compression axis (Fig. 1). The high-density fractures increase with increasing pressure, and extend throughout the matrix at 30 GPa (Fig. 2a). At 30 GPa, in addition, wider fractures form a mesh-like network throughout the matrix, dividing the sample into domains that range in size from 20 to 400 μm (Fig. 2). At 26–30 GPa, local melts also occur in the matrix, and at 34 and 36 GPa, they extend pervasively throughout the matrix, and the texture is increasingly disrupted. At 49 GPa, the matrix is totally melted.

Transmission electron microscope (TEM) observations of the

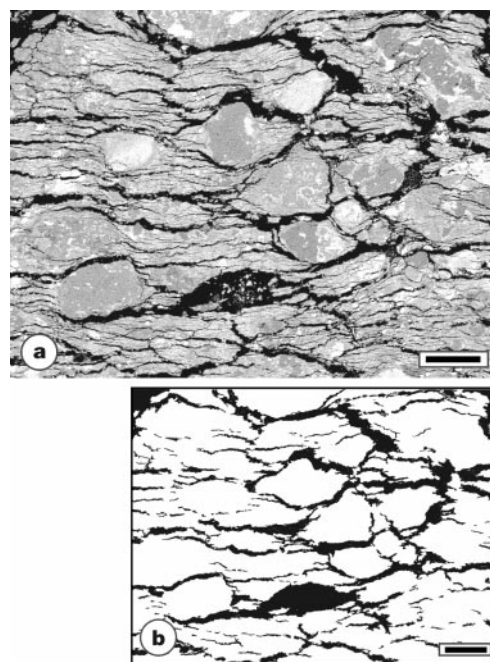


Figure 2 Back-scattered SEM images of fractures in a portion of the Murchison sample shocked at 30 GPa. **a**, In addition to narrow (1–5 μm), subparallel fractures, relatively wide (5–50 μm) fractures form a mesh-like network throughout the matrix, dividing the sample into numerous domains that range in size from 20 to 400 μm . **b**, Illustration of the mesh-like network of relatively wide (>5 μm) fractures. Scale bars, 100 μm .

particles recovered from Murchison shocked at 30 GPa reveal that they consist of abundant fine grains (0.2–1.0 μm in size) of Fe-rich olivine embedded in a Si-rich amorphous material (Fig. 3). Minor amounts of low-Ca pyroxene also occur as grains 0.2–1.0- μm in size. The Si-rich amorphous material contains numerous small rounded grains (5–20 nm) of iron sulphides and iron oxides. In places, relatively large grains (0.1–0.5 μm) of serpentine having an interlayer spacing of ~ 0.7 nm occur. However, fine-grained, fibrous serpentine (< 50 nm) and tochilinite, both of which are abundant in unshocked Murchison matrix, are absent. These observations indicate that fine-grained serpentine has been largely decomposed by post-shock heating to the Si-rich amorphous material and iron oxides, and probably also to olivine and pyroxene, and tochilinite has been decomposed to iron sulphides.

Previous studies^{7,14–16} of fine-grained, unmelted micrometeorites showed that they are typically mixtures of anhydrous and hydrous minerals and a Si-rich amorphous material in various proportions. Major minerals are olivine, low-Ca pyroxene, hydrous Mg-Fe phyllosilicates (serpentine and saponite), iron sulphides and iron oxides with various relative abundances. The Si-rich amorphous material, which is commonly intermixed with fine-grained iron sulphides and iron oxides, fills interspaces between olivine and pyroxene grains. The majority of micrometeorites have textures and mineralogy that are consistent with most of their olivine, pyroxene, iron oxides and Si-rich amorphous material having been formed by thermal transformation of phyllosilicates^{7,14,15}. These and our results indicate that the overall texture and mineralogy of the micrometeorites and the shocked Murchison matrix are closely similar to each other.

In the matrix of Allende shocked at 27–37 GPa, relatively wide (5–50 μm) fractures form sparsely, many of which are long, straight and run oblique to the compression axis. High-density fractures are not observed (Fig. 4). Minor amounts of local melts occur in the samples shocked at 27 and 37 GPa. At 41 GPa, high-density, narrow, subparallel fractures form locally in the matrix, in directions roughly perpendicular to the compression axis; but the volume fraction of fractured portions in the matrix is only $\sim 10\%$, and no mesh-like network of wide fractures is observed. At 49 GPa, the matrix is totally melted.

Such high-density fractures produced in the matrix of Murchison shocked at 26–30 GPa (Figs 1 and 2a) have never been known from natural chondritic meteorites. The sudden increase of fractures at

~ 25 GPa was probably caused by an increase of expansive force after completion of matrix pore collapse. At 21–36 GPa, local melting occurs, which indicates that shock heating also increases dramatically, simultaneously with the reduction to minute particles (comminution); the peak temperature must locally exceed 1,350 $^{\circ}\text{C}$, the melting temperature of the Murchison matrix¹⁷. The strong increase in the amount of post-shock heat is an intrinsic characteristic of porous materials¹⁸. In porous material, the kinetic energy of the projectile is partitioned more effectively into heating as well as comminution, and crushing the grains to fill the pore spaces, thus stresses of shock waves attenuate more rapidly than in nonporous material (for example, see refs 18, 19). The heating must cause dehydration of hydrous minerals, evaporation of H_2O , and increase in gas volume, and as a result, contribute to generation of great expansive force on pressure release.

Such high-density fractures have not been found from shock experiments of an ordinary chondrite at pressures even up to 60 GPa (ref. 20); ordinary chondrites represent the most abundant meteorite type ($\sim 79\%$ of meteorite falls⁸); they are anhydrous and much less porous than the carbonaceous chondrites. Our study reveals that Allende produces high-density fractures in the matrix by shock, but at pressures much higher and in a narrower range than Murchison (approximately 40–49 GPa versus 25–49 GPa). Compared to Allende, Murchison must respond to shock with a higher degree of compression owing to the higher volume proportion of porous matrix and the mechanically weak nature of fibrous phyllosilicates, and subsequently with a higher degree of expansion due to generation of great expansive force. These shock responses of Murchison would facilitate production of dense fractures in the matrix. Therefore, we conclude that the formation of high-density fractures in matrix is probably characteristic of hydrated, porous meteorites such as CI and CM chondrites.

From these results, we suggest that if hydrated, porous asteroids are shocked on the surface at pressures higher than 25 GPa, the shocked materials would be densely comminuted and/or melted, and on pressure release, the comminuted and melted particles would be explosively dispersed as dust into interplanetary space. Reflectance spectroscopy measurements^{21,22} indicate that 30–40% of known asteroids are hydrated CI-CM-like, possibly parent bodies of CI-CM chondrites, and the rest are anhydrous. Our results indicate that in the case of anhydrous asteroids, much higher shock pressures (≥ 40 GPa) would be required to comminute shocked materials, and

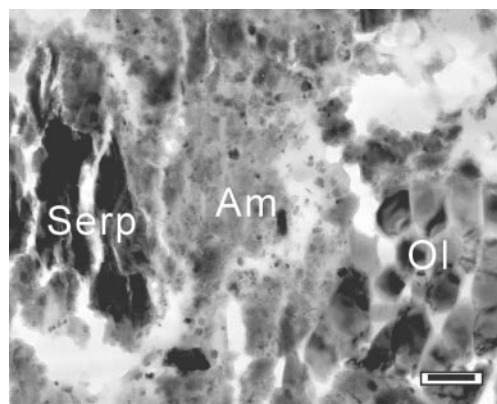


Figure 3 TEM image of the matrix of Murchison shocked at 30 GPa. Relatively large grains of Fe-rich olivine (Ol) and serpentine (Serp) are embedded in a Si-rich amorphous material (Am) that contains numerous small grains of iron sulphides and iron oxides (small grains with high contrast). Scale bar, 300 nm. Matrix grains selected from the shocked sample were thin-sectioned using an ultramicrotome. Imaging, electron diffraction and analysis were performed with a TEM (200 keV) equipped with an EDS.

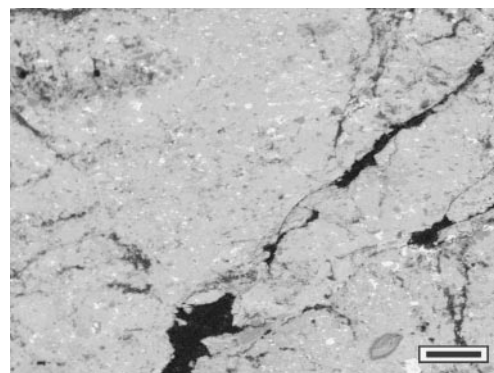


Figure 4 Back-scattered SEM image (at the same magnification as in Fig. 2a) of a portion of the matrix of Allende shocked at 37 GPa. There are no dense fractures but sparse, relatively wide fractures. Scale bar, 100 μm . The Allende samples were shocked in five experiments at peak pressures 27, 31, 37, 41 and 49 GPa. The target samples are disks of meteorites that are 10.0 mm in diameter and 1.9–3.0 mm in thickness. The velocities of projectiles range from 1.23 to 1.76 km s^{-1} .

because of generation of relatively low expansive force on pressure release, dispersion of the comminuted particles would be less effective than in the case of hydrated asteroids. Moreover, if asteroids are extremely porous and anhydrous, impact may cause compaction cratering rather than excavation²³. As a result of these differences in shock response, CI-CM-like material would become the predominant kind of dust particles produced by mutual collisions of asteroids.

Scott *et al.*²⁴ found, using petrographic observations, that almost all CM chondrites are virtually unshocked, whereas most other types of carbonaceous and ordinary chondrites, except CI type, are shocked to various pressures from 5 to 90 GPa (refs 24, 25). To explain the difference in shock record, they hypothesized that the more porous and volatile-rich CM chondrites shocked to high pressures escape from the parent asteroids more easily owing to dispersion on pressure release and these chondrites form particles that are too small to survive as meteorites. The results of our shock experiments support their hypothesis, and further suggest that CI-CM-like fragments larger than meteorite-size are not very abundant in interplanetary space, and the rate at which the atmosphere filters CI-CM meteorite infalls to the Earth is not as substantial as previously envisaged^{9,26,27}. From these results, we conclude that the difference in ejection mechanism between hydrated and anhydrous asteroids is a major factor controlling the different relative abundances of hydrated and anhydrous material in micrometeorites and meteorites.

The matrix materials in Murchison and Allende are partially melted by post-shock heating at ≥ 21 GPa and ≥ 27 GPa, respectively, and are both completely melted at 49 GPa. An ordinary chondrite exhibits extremely minor melting (~ 2 vol.% of the meteorite) even by shock at 60 GPa (ref. 20). These results indicate that the greater porosity of the carbonaceous chondrites causes higher post-shock temperatures and melting at lower pressures. Previous studies^{3–7} indicate that most micrometeorites have been significantly altered or melted by heating. The effects of heating have previously been mainly ascribed to aerodynamic drag during atmospheric entry, and the degree of heating has been thought to be related to the entry velocities of particles (and hence used to infer their orbits and sources)^{4,15,17,28,29}. However, if our model is valid, the effects of heating should also be ascribed to impact-induced shock on their parent bodies, and hence previous interpretations regarding the atmospheric entry of dust proposed from the studies of micrometeorites and interplanetary dust particles may need significant re-evaluation. □

Received 8 November 2002; accepted 4 March 2003; doi:10.1038/nature01567.

1. Love, S. G. & Brownlee, D. E. A direct measurement of the terrestrial mass accretion rate of cosmic dust. *Science* **262**, 550–553 (1993).
2. Taylor, S., Lever, J. H. & Harvey, R. P. Accretion rate of cosmic spherules measured at the South Pole. *Nature* **392**, 899–903 (1998).
3. Kurat, G., Koerber, C., Presper, T., Brandstätter, F. & Maurette, M. Petrology and geochemistry of Antarctic micrometeorites. *Geochim. Cosmochim. Acta* **58**, 3879–3904 (1994).
4. Brownlee, D. E., Bates, B. & Schramm, L. The elemental composition of stony cosmic spherules. *Meteorit. Planet. Sci.* **32**, 157–175 (1997).
5. Genge, M. J., Grady, M. M. & Hutchison, R. The textures and compositions of fine-grained Antarctic micrometeorites: Implications for comparisons with meteorites. *Geochim. Cosmochim. Acta* **61**, 5149–5162 (1997).
6. Engrand, C. & Maurette, M. Carbonaceous micrometeorites from Antarctica. *Meteorit. Planet. Sci.* **33**, 565–580 (1998).
7. Nakamura, T., Noguchi, T., Yada, T., Nakamura, Y. & Takaoka, N. Bulk mineralogy of individual micrometeorites determined by X-ray diffraction analysis and transmission electron microscopy. *Geochim. Cosmochim. Acta* **65**, 4385–4397 (2001).
8. Sears, D. W. G. & Dodd, R. T. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Mathews, M. S.) 3–31 (Univ. Arizona Press, Tucson, 1988).
9. Baldwin, B. & Sheaffer, Y. Ablation and breakup of large meteoroids during atmospheric entry. *J. Geophys. Res.* **76**, 4653–4668 (1971).
10. McSween, H. Y. Alteration in CM carbonaceous chondrites inferred from modal and chemical variations in matrix. *Geochim. Cosmochim. Acta* **43**, 1761–1770 (1979).
11. McSween, H. Y. Petrographic variations among carbonaceous chondrites of the Vigarano type. *Geochim. Cosmochim. Acta* **41**, 1777–1790 (1977).
12. Fuchs, L. H., Olsen, E. & Jensen, K. J. Mineralogy, mineral chemistry, and composition of the Murchison (C2) meteorite. *Smithson. Contrib. Earth Sci.* **10**, 1–39 (1973).

13. Corrigan, C. M. *et al.* The porosity and permeability of chondritic meteorites and interplanetary dust particles. *Meteorit. Planet. Sci.* **32**, 509–515 (1997).
14. Noguchi, T. & Nakamura, T. Mineralogy of Antarctic micrometeorites recovered from the Dome Fuji Station. *Antarct. Meteorite Res.* **13**, 285–301 (2000).
15. Greshake, A. *et al.* Heating experiments simulating atmospheric entry heating of micrometeorites: clues to their parent body sources. *Meteorit. Planet. Sci.* **33**, 267–290 (1998).
16. Jessberger, E. K. *et al.* in *Interplanetary Dust* (eds Grün, E., Gustafson, B. A. S., Dermott, S. F. & Feghtig, H.) 253–294 (Springer, Heidelberg, 2001).
17. Toppani, A., Libourel, G., Engrand, C. & Maurette, M. Experimental simulation of atmospheric entry of micrometeorites. *Meteorit. Planet. Sci.* **36**, 1377–1396 (2001).
18. Schaal, R. B., Hörz, F., Thompson, T. D. & Bauer, J. F. Shock metamorphism of granulated lunar basalt. *Proc. Lunar Planet. Sci. Conf.* **10**, 2547–2571 (1979).
19. Kieffer, S. W. Shock metamorphism of the Coconino sandstone at Meteor Crater, Arizona. *J. Geophys. Res.* **76**, 5449–5473 (1971).
20. Schmitt, R. T. Shock experiments with the H6 chondrite Kernouvé: Pressure calibration of microscopic shock effects. *Meteorit. Planet. Sci.* **35**, 545–560 (2000).
21. Gaffey, M. J., Bell, J. F. & Cruikshank, D. P. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 98–127 (Univ. Arizona Press, Tucson, 1989).
22. Jones, T. D., Lebofsky, L. A., Lewis, J. S. & Marley, M. S. The composition and origin of the C, P, and D asteroids: water as a tracer of thermal evolution in the outer belt. *Icarus* **88**, 172–192 (1990).
23. Housen, K. R., Holsapple, K. A. & Voss, M. E. Compaction as the origin of the unusual craters on the asteroid Mathilde. *Nature* **402**, 155–157 (1999).
24. Scott, E. R. D., Keil, K. & Stöffler, D. Shock metamorphism of carbonaceous chondrites. *Geochim. Cosmochim. Acta* **56**, 4281–4293 (1992).
25. Stöffler, D., Keil, K. & Scott, E. R. D. Shock metamorphism of ordinary chondrites. *Geochim. Cosmochim. Acta* **55**, 3845–3967 (1991).
26. Zolensky, M. E., Weisberg, M. K., Buchanan, P. C. & Mittlefehldt, D. W. Mineralogy of carbonaceous chondrite clasts in HED achondrites and the Moon. *Meteorit. Planet. Sci.* **31**, 518–537 (1996).
27. Sears, D. W. G. The case for chondrules and CAI being rare in the early solar system: Some implications for astrophysical models (abstract). *Lunar Planet. Sci.* **28**, 1273–1274 (1997).
28. Flynn, G. J. Atmospheric entry heating: A criterion to distinguish between asteroidal and cometary sources of interplanetary dust. *Icarus* **77**, 287–310 (1989).
29. Love, S. G. & Brownlee, D. E. Heating and thermal transformation of micrometeoroids entering the Earth's atmosphere. *Icarus* **89**, 26–43 (1991).
30. Tomeoka, K., Yamahana, Y. & Sekine, T. Experimental shock metamorphism of the Murchison CM carbonaceous chondrite. *Geochim. Cosmochim. Acta* **63**, 3683–3703 (1999).

Acknowledgements We thank T. Hiroi, L. P. Keller and T. Mukai for discussions. We acknowledge support by a Grant-in-Aid from the Ministry of Education, Science and Culture, Japan, and a grant from the Japan Society for the Promotion of Science.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.T. (tomeoka@kobe-u.ac.jp).

Bose–Einstein condensation of the triplet states in the magnetic insulator TiCuCl_3

Ch. Rüegg*, N. Cavadini*, A. Furrer*, H.-U. Güdel†, K. Krämer‡, H. Mutka‡, A. Wildes‡, K. Habicht§ & P. Vorderwisch§

* Laboratory for Neutron Scattering, ETH Zürich & Paul Scherrer Institut, 5232 Villigen PSI, Switzerland

† Department of Chemistry and Biochemistry, Universität Bern, 3000 Bern 9, Switzerland

‡ Institut Laue-Langevin, BP 156, 38042 Grenoble Cedex 9, France

§ BENSC, Hahn-Meitner-Institut, 14109 Berlin Wannsee, Germany

Bose–Einstein condensation denotes the formation of a collective quantum ground state of identical particles with integer spin or intrinsic angular momentum. In magnetic insulators, the magnetic properties are due to the unpaired shell electrons that have half-integer spin. However, in some such compounds (KCuCl_3 and TiCuCl_3), two Cu^{2+} ions are antiferromagnetically coupled¹ to form a dimer in a crystalline network: the dimer ground state is a spin singlet (total spin zero), separated by an energy gap from the excited triplet state (total spin one). In these dimer compounds, Bose–Einstein condensation becomes theoretically

possible². At a critical external magnetic field, the energy of one of the Zeeman split triplet components (a type of boson) intersects the ground-state singlet, resulting in long-range magnetic order; this transition represents a quantum critical point at which Bose–Einstein condensation occurs. Here we report an experimental investigation of the excitation spectrum in such a field-induced magnetically ordered state, using inelastic neutron scattering measurements of TlCuCl_3 single crystals. We verify unambiguously the theoretically predicted³ gapless Goldstone mode characteristic of the Bose–Einstein condensation of the triplet states.

One essential parameter for classification of particles is their spin or intrinsic angular momentum. Half-integer-spin fermions are constrained by the Pauli exclusion principle, whereas integer-spin bosons are not. As a result, the spin determines the nature of the energy distribution in a collection of the particles: upon cooling, fermions create a ‘Fermi sea’ whereas bosons create a ‘condensate’. In 1924, Bose and Einstein pointed out that bosons could condense in unlimited numbers into a single ground state, as they are not constrained by the Pauli exclusion principle^{4,5}. The collection into a single state is therefore called Bose–Einstein condensation (BEC). Little notice was taken of this curious possibility until the anomalous behaviour of liquid helium was studied carefully⁶. When ^4He is cooled to a critical temperature of 2.17 K, a remarkable discontinuity in heat capacity occurs, the liquid density drops, and a fraction of the liquid becomes a zero-viscosity superfluid. Superfluidity arises from the fraction of helium atoms that has condensed to the lowest possible energy. A condensation effect is also credited with producing superconductivity. In the Bardeen–Cooper–Schrieffer theory, pairs of electrons are coupled by lattice interactions: the pairs (called Cooper pairs) act like bosons, and can condense into a state of zero electrical resistance⁷. Finally, in 1995, alkali atoms were demonstrated to undergo BEC, and thereby to form a novel state of matter. Magnetically trapped, spin-polarized diluted gases of alkali atoms were cooled down to the nanokelvin temperature scale, where a fraction of the individual atoms condense into a ‘superatom’ behaving as a single entity⁸.

Magnetic insulators provide an unrivalled experimental testing ground for the understanding of collective quantum phenomena. This relates both to the description of the magnetic degrees of freedom in terms of a well-defined spin hamiltonian, as well as the applicability of microscopic techniques like nuclear magnetic resonance or inelastic neutron scattering. BEC has so far escaped an experimental investigation within this framework. However, newly discovered materials, based on a crystalline network of dimers, provide high expectations of BEC behaviour in these insulators¹. An unusual kind of magnetic order occurs in these materials when the Zeeman energy $g\mu_B H$ (where g is the Landé factor, μ_B is the Bohr magneton, and H is the magnetic field) overcomes the spin gap between the $S = 0$ singlet ground state and the $S = 1$ excited triplet states (S denotes the quantum number for the total spin)^{9,10}. The magnetic order is ascribed to BEC by applying a mapping of the triplet states to a diluted Bose gas. Notable realization is documented in the dimer compounds KCuCl_3 and TlCuCl_3 , where the modest spin gap corresponds to the critical magnetic field $H_c \approx 23$ T and $H_c \approx 6$ T, respectively¹¹. Detailed experimental studies at $H > H_c$ have been performed on TlCuCl_3 by bulk measurements^{12,13} and neutron diffraction¹⁴. A consistent picture of the field-driven magnetically ordered phase has been obtained in terms of BEC of the triplet states^{2,3}. The condensate originates from the coherent superposition of the singlet and the $S_z = +1$ triplet components throughout the dimer network (S_z denotes the quantum number for the spin component along the quantization axis). Dynamic probes like inelastic neutron scattering (INS) provide valuable tools for its further characterization.

Until now, the magnetic excited states in TlCuCl_3 have been investigated by INS below the critical magnetic field $H_c \approx 6$ T. The

energy minimum of the triplet modes at $H = 0$ T has been identified with the magnetic Bragg points progressively emerging at $H > H_c$ (refs 15, 16). The Zeeman splitting of the degenerate triplet modes $S_z = +1, 0, -1$ has been further demonstrated at $0 < H < H_c$ (ref. 17). These results fully characterized the magnetic excitations in the spin-liquid state, and confirmed the picture of the soft Zeeman mode driving the quantum phase transition at $H = H_c$ from the spin-liquid state to the field-induced magnetically ordered state. The evolution of the magnetic properties across H_c is apt to demonstrate a quantum phase transition, which occurs as a function of the magnetic field at fixed temperature ‘ $T = 0$ K’ (ref. 18). But experimental INS observations directly at $H \approx H_c$ are generally limited by the finite instrumental resolution and the elastic nuclear incoherent scattering¹⁷. Dynamic predictions for the magnetic excitations at $H > H_c$ have been recently elaborated in the context of the BEC of the triplet states³. In particular, the condensate was postulated to support a gapless linear mode, which corresponds to the Goldstone mode of the ordered phase. No INS investigation of the field-driven ordered phase of dimer spin systems is available in the literature. However, besides TlCuCl_3 , several magnetic insulators have been recently characterized by INS with related aims—for example, $\text{SrCu}_2(\text{BO}_3)_2$ (ref. 19), $\text{Cu}(\text{NO}_3)_2 \cdot 2.5\text{D}_2\text{O}$ (ref. 20) and $\text{Cu}_2(\text{C}_5\text{H}_{12}\text{N}_2)_2\text{Cl}_4$, known as CuHpCl (ref. 21). The possibility of condensing elementary magnetic states by applying an external magnetic field has further been proposed for the one-dimensional case of $S = 1$ Haldane spin gap chains in the high-field phase^{22,23}. The compounds $\text{Ni}(\text{C}_2\text{H}_8\text{N}_2)_2\text{NO}_2\text{ClO}_4$ and $\text{Ni}(\text{C}_5\text{H}_{14}\text{N}_2)_2\text{N}_3(\text{PF}_6)$, known as NENP and NDMAP, respectively, clearly show the Haldane gap at zero field, but are unfortunately dominated by the presence of a staggered g -tensor or a single-ion anisotropy causing the spin gap to reopen above H_c (refs 24, 25). There is no evidence for such mechanisms in TlCuCl_3 (refs 12, 14, 17).

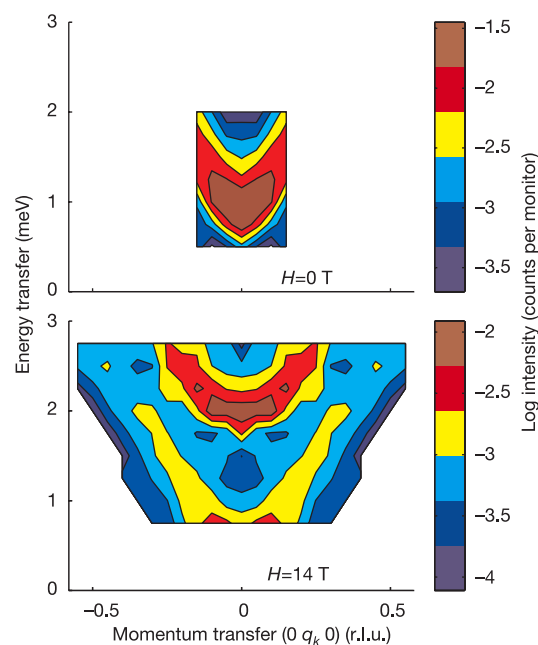


Figure 1 Contour plot of the magnetic inelastic neutron scattering intensity measured in TlCuCl_3 . Data are shown for an external magnetic field $H = 0$ T (spin-liquid state) and $H = 14$ T (field-induced ordered state), for fixed $T = 1.5$ K. The x -coordinate shows the wavevector transfer $\mathbf{Q} = \mathbf{q} + \boldsymbol{\tau}$ in reciprocal lattice units (r.l.u.) relative to the Bragg point $\boldsymbol{\tau} = (0\ 4\ 0)$ r.l.u. The y -coordinate shows the energy transfer in meV. The colours indicate the neutron scattering intensity normalized to give counts per monitor. The data are merged in both cases according to a bin size of $0.05 \cdot 0.25$ (r.l.u.) $\cdot$ (meV) to yield a continuous colour map, and are symmetrized around $\mathbf{q} = 0$. Comments are provided in the text.

Here we report the INS investigation of the field-driven ordered phase in TiCuCl_3 , measured at $H > H_c$ up to $H = 14$ T. We unambiguously demonstrate the coexistence of a novel low-lying mode and two higher-lying modes from the neutron spectra measured around the magnetic Bragg point at $H = 14$ T. We further demonstrate the gapless and linear nature of the low-lying mode, compared with the gapped and quadratic nature of the two higher-lying modes. The measurement of the spectral evolution at intermediate fields $0 < H < 14$ T allows an intuitive understanding of our experimental observations. The two higher-lying modes correspond to the renormalized $S_z = 0$ and $S_z = -1$ Zeeman modes, which approximately scale for $H > H_c$ as $g\mu_B H$ and $2g\mu_B H$, respectively, according to the Zeeman level-crossing within a dimer scheme²⁶. The low-lying mode corresponds to the counterpart of the $S_z = +1$ Zeeman mode, which drives the field-induced order at the Bragg point but remains gapless for $H > H_c$, in contradiction to the Zeeman level-crossing.

In Fig. 1, the neutron spectra observed in TiCuCl_3 at $H = 0$ T (top) and $H = 14$ T (bottom) for fixed $T = 1.5$ K are compared, as measured under the same instrumental conditions during the same session. A substantial portion of the $(\mathbf{Q}, \omega)$ -space around the Bragg point is covered ($\mathbf{Q}$ is the wavevector transfer and $E = \hbar\omega$ is the energy transfer). Coexistence of low-lying excitations and higher-lying excitations is clearly seen in the field-induced ordered phase at $H = 14$ T. Strong field-induced magnetic Bragg contamination prevented further data collection below 0.75 meV (white area). In Fig. 2, the evolution of the magnetic excitations is summarized below, across and above H_c for fixed $T = 1.5$ K (red points) and $T = 50$ mK (blue points), as measured by energy scans at the Bragg point. Symbols correspond to the energy determined from resolution-limited fits to the observed neutron profiles, with full consideration of the instrumental convolution. Fit tolerances denoted by vertical error bars are of the size of the symbols. At $H > H_c$, we notice both the lack of the $S_z = +1$ Zeeman mode at finite energies (which remains gapless), as well as the characteristic renormalization of the $S_z = 0$ and $S_z = -1$ Zeeman modes, as anticipated. In

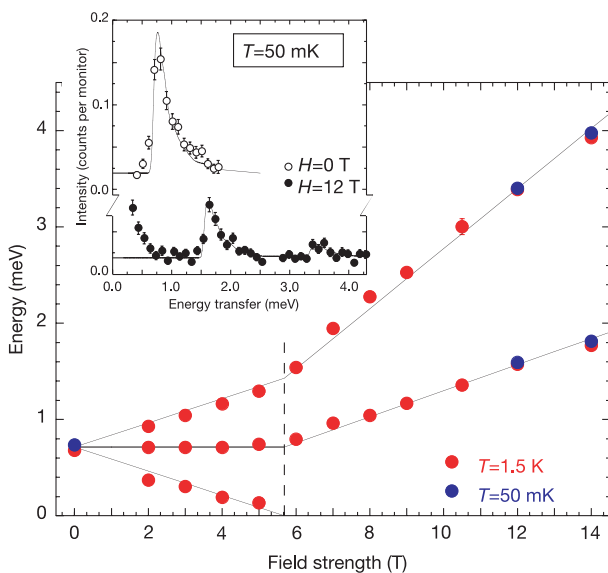


Figure 2 Dependence on the external magnetic field H of the magnetic excitation energy measured in TiCuCl_3 at the Bragg point $\tau = (0\ 4\ 0)$ r.l.u. Data are shown for fixed $T = 1.5$ K (red symbols) and $T = 50$ mK (blue symbols). Data are extracted from least-squares fits to the neutron scattering spectra (inset), curves reflect a Zeeman model. The quantum critical field $H = H_c$ is denoted by the dashed boundary. Comments and interpretations are given in the text.

Fig. 3, the energy dispersion of the low-lying mode around the Bragg point is presented at $H = 14$ T for fixed $T = 1.5$ K (red points) and $T = 50$ mK (blue points), as measured by $\mathbf{Q}$ -scans at fixed energy transfer. Symbols correspond to the wavevector determined from resolution-limited fits to the observed neutron profiles. Fit tolerances are the same size as the symbols. The origin of the low-lying mode extrapolates to the Bragg point from the linear fits as shown. Within statistical uncertainty, our quantitative conclusions on the higher-lying modes (Fig. 2) and on the low-lying mode (Fig. 3) apply to the data sets at both $T = 1.5$ K and $T = 50$ mK. This supports the robustness of the field-induced magnetically ordered phase in TiCuCl_3 , which at $H = 14$ T is observed below $T = 8$ K (ref. 12).

We have described above the magnetic excitation spectrum realized in the field-driven ordered phase of a dimer compound. This phase is theoretically interpreted as BEC within a framework of elementary magnetic states^{2,3}. Our experimental observations are in accordance with the dynamical predictions³, which are confirmed here for the complete excitation spectrum. Modest quantitative discrepancy, limited to the stiffness of the gapless mode, is observed against the numerical calculations from the same theory, with no free parameters with respect to $H = 0$ T (M. Matsumoto, personal communication). We briefly comment on the origin of this discrepancy. One possible origin lies in the simplification of the interaction parameters used for the description of the triplet dispersion in the spin-liquid phase, and fixed for the subsequent calculations in the field-induced ordered phase³. Corrections of higher order appropriate to the dimer network of KCuCl_3 and TiCuCl_3 are available²⁷, but the extraction of the complete single-ion interactions from the dimer interactions remains a formidable task if restricted to the spin-liquid state at $H < H_c$. In a recent study of the high-field magnetization curve of KCuCl_3 , the former were however addressed by the measurement of the magnetic saturation field²⁸. Unfortunately, a comparative study of TiCuCl_3 is lacking owing to the excessive value of the magnetic saturation field^{3,11}. A second possible origin of the discrepancy lies in the neglect of magneto-elastic effects connected with the field-induced ordered phase. Spin-phonon coupling is explicitly mentioned in the recent theoretical study of the electronic structure of TiCuCl_3 (ref. 29). Direct experimental evidence is further claimed from sound-wave propagation measurements in related NH_4CuCl_3 (ref. 30). A larger INS survey, extended to different directions in reciprocal space, is

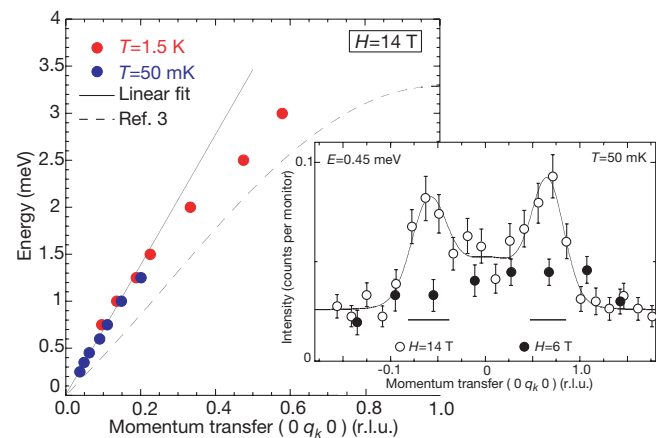


Figure 3 Energy dispersion of the low-lying magnetic excitations measured in TiCuCl_3 at $H = 14$ T. Data are shown for fixed $T = 1.5$ K (red symbols) and $T = 50$ mK (blue symbols). The x - and y -coordinates are as in Fig. 1. Data are extracted from least-squares fits to the neutron scattering spectra (inset); the continuous curve is a fit to a linear energy dispersion, the dashed curve is a theoretical expectation with no disposable parameters. Comments and interpretations are given in the text.

needed to appropriately discriminate between these possibilities in the present case.

We have investigated by inelastic neutron scattering the magnetic excitations in the field-driven ordered phase of TiCuCl_3 . Our main result is the evidence of a gapless mode emerging at the field-induced magnetic Bragg point. The gapless mode is accompanied by two renormalized Zeeman modes at higher energies. We find almost complete agreement between experimental observations and theoretical predictions that are based on a BEC theory. However, this theory does not involve a dilute gas cooled below the de Broglie temperature, but rather dimer states split at high magnetic fields. We hope that our approach to the study of a fundamental quantum effect like BEC will enable further theoretical and experimental work. □

Methods

The inelastic neutron scattering investigations up to $H = 14$ T were performed on the three-axis spectrometer V2 at the Hahn-Meitner-Institut, Germany. Preliminary measurements up to $H = 6$ T were performed on the three-axis spectrometers IN14 at the Institut Laue-Langevin, France, and on TASP at the spallation neutron source SINQ, Switzerland. All instruments featured initial and final energy selection by pyrolytic graphite (0 0 2) crystals. The measurements were performed at fixed final energy, with a cooled beryllium filter in front of the analyser set at 4.7 meV. Standard instrumental configurations with 40° -open-open and 60° - 60° - 60° horizontal collimation were adopted. The typical elastic energy resolution was 0.2 meV (full-width at half-maximum). Superconducting cryomagnets with split coils were adopted on all instruments. The cooling of the sample in the cryomagnets was achieved by ^4He flow through a root pump ($T = 1.5$ K) and by a ^3He - ^4He dilution insert system ($T = 50$ mK), producing the data set presented in the text. Further details of the sample environment will be provided elsewhere.

Received 24 December 2002; accepted 17 March 2003; doi:10.1038/nature01617.

- Rice, T. M. To condense or not to condense. *Science* **298**, 760–761 (2002).
- Nikuni, T., Oshikawa, M., Oosawa, A. & Tanaka, H. Bose–Einstein condensation of diluted magnons in TiCuCl_3 . *Phys. Rev. Lett.* **84**, 5868–5871 (2000).
- Matsumoto, M., Normand, B., Rice, T. M. & Sigrist, M. Magnon dispersion in the field-induced magnetically ordered phase of TiCuCl_3 . *Phys. Rev. Lett.* **89**, 077203 (2002).
- Bose, S. N. Plancks Gesetz und Lichtquantenhypothese. *Z. Phys.* **26**, 178–181 (1924).
- Einstein, A. Quantentheorie des einatomigen idealen Gases. *Sitzungsber. Kgl. Preuss. Akad. Wiss.* 261–267 (1924).
- London, F. The λ -phenomenon of liquid helium and the Bose–Einstein degeneracy. *Nature* **141**, 643–644 (1938).
- Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
- Anglin, J. R. & Ketterle, W. Bose–Einstein condensation of atomic gases. *Nature* **416**, 211–218 (2002).
- Giamarchi, T. & Tsvetlik, A. M. Coupled ladders in a magnetic field. *Phys. Rev. B* **59**, 11398–11407 (1999).
- Wessel, S. & Haas, S. Three-dimensional ordering in weakly coupled antiferromagnetic ladders and chains. *Phys. Rev. B* **62**, 316–323 (2000).
- Shiramura, W. *et al.* High-field magnetization processes of double spin chain systems KCuCl_3 and TiCuCl_3 . *J. Phys. Soc. Jpn* **66**, 1900–1903 (1997).
- Oosawa, A., Ishii, M. & Tanaka, H. Field-induced three-dimensional magnetic ordering in the spin gap system TiCuCl_3 . *J. Phys. Condens. Matter* **11**, 265–271 (1999).
- Oosawa, A., Katori, H. A. & Tanaka, H. Specific heat study of the field-induced magnetic ordering in the spin gap system TiCuCl_3 . *Phys. Rev. B* **63**, 134416 (2001).
- Tanaka, H. *et al.* Observation of field-induced transverse Néel ordering in the spin gap system TiCuCl_3 . *J. Phys. Soc. Jpn* **70**, 939–942 (2001).
- Cavadini, N. *et al.* Magnetic excitations in the quantum spin system TiCuCl_3 . *Phys. Rev. B* **63**, 172414 (2001).
- Oosawa, A. *et al.* Magnetic excitations in the spin gap system TiCuCl_3 . *Phys. Rev. B* **65**, 094426 (2002).
- Cavadini, N. *et al.* Triplet excitations in low- H_c spin gap systems KCuCl_3 and TiCuCl_3 : an inelastic neutron scattering study. *Phys. Rev. B* **65**, 132415 (2002).
- Sachdev, S. Quantum criticality: competing ground states in low dimensions. *Science* **288**, 475–480 (2000).
- Kageyama, H. *et al.* Direct evidence for the localized singlet-triplet excitations and the dispersive multitriplet excitations in $\text{SrCu}_2(\text{BO}_3)_2$. *Phys. Rev. Lett.* **84**, 5876–5879 (2000).
- Xu, G. Y., Broholm, C., Reich, D. H. & Adams, M. A. Triplet waves in a quantum spin liquid. *Phys. Rev. Lett.* **84**, 4465–4468 (2000).
- Stone, M. B., Zaliznyak, I., Reich, D. H. & Broholm, C. Frustrated three-dimensional quantum spin liquid in CuHpCl . *Phys. Rev. B* **65**, 064423 (2002).
- Affleck, I. Theory of Haldane-gap antiferromagnets in applied fields. *Phys. Rev. B* **41**, 6697–6702 (1990).
- Sachdev, S., Senthil, T. & Shankar, R. Finite-temperature properties of quantum antiferromagnets in a uniform magnetic field in one and two dimensions. *Phys. Rev. B* **50**, 258–272 (1994).
- Enderle, M. *et al.* High-field spin dynamics of antiferromagnetic quantum spin chains. *Physica B* **276–278**, 560–561 (2000).
- Zheludev, A. *et al.* Massive triplet excitations in a magnetized anisotropic Haldane spin chain. *Phys. Rev. Lett.* (submitted).

- Rüegg, Ch. *et al.* Spin dynamics in the high field phase of quantum critical $S = 1/2$ TiCuCl_3 . *Appl. Phys. A* **74**, S840–S842 (2002).
- Müller, M. & Mikeska, H. J. On the dynamics of coupled $S = 1/2$ antiferromagnetic zigzag chains. *J. Phys. Condens. Matter* **12**, 7633–7645 (2000).
- Oosawa, A. *et al.* Field-induced magnetic ordering in the quantum spin system KCuCl_3 . *Phys. Rev. B* **66**, 104405 (2002).
- Saha-Dasgupta, T. & Valenti, R. Comparative study between two quantum spin systems KCuCl_3 and TiCuCl_3 . *Europhys. Lett.* **60**, 309–315 (2002).
- Schmidt, S. *et al.* Phonon effects and ESR in NH_4CuCl_3 . *Europhys. Lett.* **53**, 591–597 (2001).

Acknowledgements We thank M. Matsumoto, B. Normand, T. M. Rice, M. Sigrist, M. Müller and H.-J. Mikeska for discussions, M. Meissner, S. Kausche and P. Smeibidl for assistance during the V2 measurements, and F. Thomas for assistance during the IN14 measurements. This work was supported partially by the Swiss National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R. (christian.rueegg@psi.ch).

The origin of multiple superconducting gaps in MgB_2

S. Souma*, Y. Machida†, T. Sato*, T. Takahashi*, H. Matsui*
S.-C. Wang‡, H. Ding‡, A. Kaminski§, J. C. Campuzano§, S. Sasaki†
& K. Kadowaki¶

* Department of Physics, Tohoku University, Sendai 980-8578, Japan

† Materials and Structures Laboratory, Tokyo Institute of Technology, Yokohama 226-8503, Japan

‡ Department of Physics, Boston College, Chestnut Hill, Massachusetts 02467, USA

§ Materials Sciences Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

|| Department of Physics, University of Illinois at Chicago, Chicago, Illinois 60607, USA

¶ Institute of Materials Science, University of Tsukuba, Tsukuba 305-8573, Japan

Magnesium diboride, MgB_2 , has the highest transition temperature ($T_c = 39$ K) of the known metallic superconductors¹. Whether the anomalously high T_c can be described within the conventional BCS (Bardeen–Cooper–Schrieffer) framework² has been debated. The key to understanding superconductivity lies with the ‘superconducting energy gap’ associated with the formation of the superconducting pairs. Recently, the existence of two kinds of superconducting gaps in MgB_2 has been suggested by several experiments^{3–9}; this is in contrast to both conventional and high- T_c superconductors. A clear demonstration of two gaps has not yet been made because the previous experiments lacked the ability to resolve the momentum of the superconducting electrons. Here we report direct experimental evidence for the two-band superconductivity in MgB_2 , by separately observing the superconducting gaps of the σ and π bands (as well as a surface band). The gaps have distinctly different sizes, which unambiguously establishes MgB_2 as a two-gap superconductor^{10,11}.

Soon after the discovery of MgB_2 , many experimental studies indicated that MgB_2 should be basically classified as a conventional phonon-mediated BCS superconductor^{12–16}. Recently, however, its deviation from the simple BCS framework has been inferred from several experiments using high-quality samples^{3–9,17}. The most striking deviation is the complex superconducting order parameter, referred to as the ‘multi-component superconducting gap’ or ‘multi-gap’, indicating that two or more superconducting gaps with different sizes develop simultaneously at T_c between the occupied and unoccupied electronic states^{10,11}. This multi-gap

behaviour has been analysed by several models^{10–13,18–20} based on the different role of σ and π bands^{10,11,18} or surface and bulk^{12,13,19}, or the strongly anisotropic coupling to lattice vibrations (phonons)²⁰.

Of these, the two-band model^{10,11} may be theoretically described as follows: electrons in the σ bands are strongly coupled to phonons confined within the honeycombed boron layer and give rise to a large gap, whereas a relatively small gap opens in the π band due to the weak electron–phonon coupling. There has been no direct experimental evidence, however, that establishes the two different gaps originating in the σ and π bands. Previous experiments such as the specific-heat measurements^{3,4}, and Raman^{6,7} and tunnelling^{8,9} spectroscopies provide the information averaged over all the momentum space and therefore could not separate the different bands. In contrast, the present high-resolution angle-resolved photoemission spectroscopy (ARPES) experiment has succeeded in resolving the σ and π bands and directly observing the superconducting gaps.

Figure 1a shows ARPES spectra of MgB₂ measured along the Γ KM(AHL) direction (see inset to Fig. 1a) with 28-eV photons at 45 K. We find several band dispersions which cross the Fermi level (E_F) in this momentum region. To visualize the band dispersion more clearly, we plot the ARPES intensity as a function of wave vector and binding energy in Fig. 1b. We find three distinct band dispersions. The first is a large electron-like band which crosses E_F near the K(H) point. The second is a hole-like dispersion centred at Γ (A), crossing E_F near Γ (A). The third is a small electron-like pocket centred at Γ (A).

This result is consistent with the previous ARPES report²¹. According to the band calculation²², the first and second bands are attributed to the boron $2p_\pi$ and $2p_\sigma$ bands, respectively. The rather broad feature of the π band reflects its three-dimensional nature, while the sharp σ band shows the strong two-dimensional character. The third band, which is not seen in the band calculation, has been ascribed to a surface state²¹. The observed clear difference in the E_F -crossing point among the three bands enables us to perform precise ARPES measurements on the superconducting gaps of different bands separately.

We measured ARPES spectra in the vicinity of E_F at three different points on the Fermi surfaces (points A, B and C in Fig. 1b) at two temperatures below and above T_c (17 K and 45 K). The results are shown in Fig. 2. The ARPES spectrum from the σ band shows a remarkable temperature dependence which cannot be explained by the simple temperature effect due to the Fermi–Dirac function. The midpoint of the leading edge in the spectrum at 17 K is not at E_F but is shifted by about 2 meV toward the higher binding energy, while the leading-edge midpoint at 45 K is located at E_F . This clearly indicates that a superconducting gap opens at 17 K in the σ band. We observe a small piling-up effect of spectral weight around 7–15 meV in the 17 K spectrum, which indicates the emergence of a superconducting coherent peak below T_c , as observed in other superconductors²³. We find it surprising that a similar gap-opening behaviour is observed in the surface band as seen in Fig. 2. This point will be discussed later.

In contrast to the σ and surface bands, the π band does not show a clear gap-opening behaviour. The shift of the midpoint of the leading edge at 17 K is much less (0–1 meV) and no clear piling-up effect around 10 meV is seen, suggesting that no or a very small superconducting gap opens in the π band. This reveals the clear difference in the contribution to the superconductivity among three bands, supporting the multi-band superconductivity scenario in MgB₂ (refs 10, 11).

To estimate the gap size (Δ), we numerically fitted the spectrum using the BCS spectral function with an integrated-type background representing the incoherent part of the spectral function and inelastically scattered electrons, multiplied by the Fermi–Dirac function and convoluted with the energy resolution, as has been employed in high- T_c copper oxides²⁴. As shown in Fig. 2, ARPES

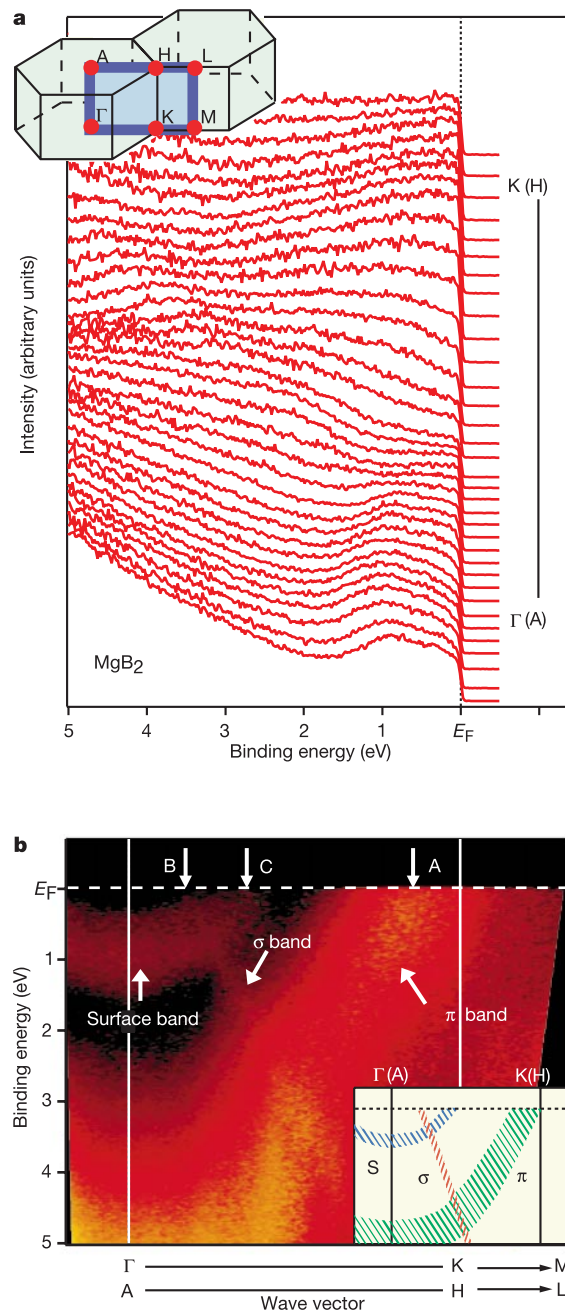


Figure 1 Experimental band structure near E_F of MgB₂ obtained by ARPES. **a**, ARPES spectra of MgB₂ measured at 45 K along the Γ KM(AHL) direction in the Brillouin zone (inset to **a**) with 28-eV photons. **b**, ARPES-intensity plot of MgB₂ as a function of wave vector and binding energy. White arrows A, B and C indicate the location of Fermi-level crossing points for the π , σ and surface bands. Inset to **b** shows the schematic view of the band dispersion; S indicates surface. Single crystals of MgB₂ were grown in a closed stainless steel tube filled with argon gas. The tube was heated up to 1,200 °C and slowly cooled down to room temperature in 12 h. Details of sample preparation are described elsewhere²⁶. Magnetization measurement confirmed that crystals exhibit superconductivity at 38 K with a transition width of $\Delta T_c = 0.8$ K. The typical size of crystals used in ARPES measurements is about $100 \times 100 \times 10 \mu\text{m}^3$. ARPES measurements were performed using a SCIENTA SES-2002 spectrometer at the undulator 4 m-NIM beam line at the Synchrotron Radiation Center, Wisconsin. The energy and angular (momentum) resolution were set to 10 meV and 0.3° ($0.01 \text{ \AA}^{-1}$), respectively. We obtained a clean surface for ARPES measurements by cleaving a single crystal *in situ* along the (0001) plane in an ultrahigh vacuum (better than 2×10^{-10} torr). The crystal orientation was determined by Laue X-ray diffraction before the ARPES measurement. Polarization vector of photons was set parallel to the Γ KM(AHL) direction.

spectra at 17 K for the σ , π and surface bands are satisfactorily fitted with gap sizes of $\Delta = 6.5 \pm 0.5$ meV, 1.5 ± 0.5 meV, and 6.0 ± 0.5 meV, respectively, with broadening factor of 1.0 meV for the σ and surface bands and 0.5 meV for the π band. A slight deviation above E_F observed for the σ and surface bands may be due to a normal metallic region at the surface²⁵.

From the viewpoint of gap size, we consider that the present ARPES experiment has observed 'two' gaps in MgB_2 because the size is almost the same between the σ and surface bands. Thus far, many experiments, in particular those using tunnelling spectroscopy, have reported the multi-gap feature in MgB_2 and the reported gap sizes are roughly categorized into two groups: small and large superconducting gaps. The small gap has a value of 1–4 meV and the large gap has a value of 5.5–10 meV. The present ARPES result shown above seems to be consistent with these previous reports.

The two-gap nature in MgB_2 reported by previous studies has been interpreted by several models: (1) the simultaneous observation of the genuine bulk superconductivity and the weakened superconductivity in the surface layer^{12,13}; (2) the proximity effect in a thin metallic surface layer¹⁹; (3) the strongly anisotropic coupling to phonons²⁰; (4) the contribution from the σ and π bands, which give the large and small gaps, respectively^{10,11}; and (5) as for model (4) but with opposite band contributions, that is, the π band produces a large gap¹⁸.

It is clear from the present ARPES results that the σ and the surface bands have a large gap of 6–7 meV whereas the π band shows a small gap of 1–2 meV. Taking into account that the superconductivity of MgB_2 is bulk in nature^{1,3}, we conclude that the σ band is dominant in the superconductivity of MgB_2 with a stronger coupling to phonons, while the π band is less important to the superconductivity with a much weaker coupling. This unambigu-

ously indicates that the two-band model^{10,11} describes the superconductivity of MgB_2 most appropriately.

We also conclude from the present ARPES results that the large gap observed by surface-sensitive techniques such as tunnelling spectroscopy may be due to both the σ and the surface bands, because the gap size is almost the same for the two bands. A large variation in the size of the large gap may stem from the various surface conditions, depending on the surface preparation method. Although why the surface band exhibits a similar gap size to the σ band is unclear at present, the proximity of the two bands in the momentum and energy spaces as seen in Fig. 1 may account for it, because the interaction between the surface and σ bands is expected to be larger than that between the surface and π bands. $\square$

Received 27 January; accepted 8 April 2003; doi:10.1038/nature01619.

1. Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y. & Akimitsu, J. Superconductivity at 39 K in magnesium diboride. *Nature* **410**, 63–64 (2001).
2. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
3. Bouquet, F., Fisher, R. A., Phillips, N. E., Hinks, D. G. & Jorgensen, J. D. Specific heat of MgB_2 : evidence for a second energy gap. *Phys. Rev. Lett.* **87**, 047001 (2001).
4. Bouquet, F. *et al.* Specific heat of single crystal MgB_2 : a two-band superconductor with two different anisotropies. *Phys. Rev. Lett.* **89**, 257001 (2002).
5. Tsuda, S. *et al.* Evidence for a multiple superconducting gap in MgB_2 from high-resolution photoemission spectroscopy. *Phys. Rev. Lett.* **87**, 177006 (2001).
6. Chen, X. K., Konstantinović, M. J., Irwin, J. C., Lawrie, D. D. & Franck, J. P. Evidence for two superconducting gaps in MgB_2 . *Phys. Rev. Lett.* **87**, 157002 (2001).
7. Quilty, J. W., Lee, S., Tajima, S. & Yamanaka, A. *c*-axis Raman scattering in MgB_2 : observation of a dirty-limit gap in the π -bands. Preprint cond-mat/0206506 at (<http://xxx.lanl.gov>) (2002).
8. Szabó, P. *et al.* Evidence for two superconducting energy gaps in MgB_2 by point-contact spectroscopy. *Phys. Rev. Lett.* **87**, 137005 (2001).
9. Gonnelli, R. S. *et al.* Direct evidence for two-band superconductivity in MgB_2 single crystals from directional point-contact spectroscopy in magnetic fields. *Phys. Rev. Lett.* **89**, 247004 (2002).
10. Liu, A. Y., Mazin, I. I. & Kortus, J. Beyond Eliashberg superconductivity in MgB_2 : anharmonicity, two-phonon scattering, and multiple gaps. *Phys. Rev. Lett.* **87**, 087005 (2001).
11. Choi, H. J., Roundy, D., Sun, H., Cohen, M. L. & Louie, S. G. The origin of the anomalous superconducting properties of MgB_2 . *Nature* **418**, 758–760 (2002).
12. Karapetrov, G., Iavarone, M., Kwok, W. K., Crabtree, G. W. & Hinks, D. G. Scanning tunneling spectroscopy in MgB_2 . *Phys. Rev. Lett.* **86**, 4374–4377 (2001).
13. Schmidt, H., Zasadzinski, J. F., Gray, K. E. & Hinks, D. G. Energy gap from tunneling and metallic contacts onto MgB_2 : possible evidence for a weakened surface layer. *Phys. Rev. B* **63**, 220504 (2001).
14. Takahashi, T., Sato, T., Souma, S., Muranaka, T. & Akimitsu, J. High-resolution photoemission study of MgB_2 . *Phys. Rev. Lett.* **86**, 4915–4917 (2001).
15. Bud'ko, S. L. *et al.* Boron isotope effect in superconducting MgB_2 . *Phys. Rev. Lett.* **86**, 1877–1880 (2001).
16. Kotegawa, H., Ishida, K., Kitaoka, Y., Muranaka, T. & Akimitsu, J. Evidence for strong-coupling *s*-wave superconductivity in MgB_2 : ^{11}B NMR study. *Phys. Rev. Lett.* **87**, 127001 (2001).
17. Buzea, C. & Yamashita, T. Review of the superconducting properties of MgB_2 . *Supercond. Sci. Technol.* **14**, R115–R146 (2001).
18. Joas, C., Eremin, I., Manske, D. & Bennemann, K. H. Theory for phonon-induced superconductivity in MgB_2 . *Phys. Rev. B* **65**, 132518 (2002).
19. McMillan, W. L. Tunneling model of the superconducting proximity effect. *Phys. Rev.* **175**, 537–542 (1968).
20. Seneor, P. *et al.* Spectroscopic evidence for anisotropic *s*-wave pairing symmetry in MgB_2 . *Phys. Rev. B* **65**, 012505 (2001).
21. Uchiyama, H. *et al.* Electronic structure of MgB_2 from angle-resolved photoemission spectroscopy. *Phys. Rev. Lett.* **88**, 157002 (2002).
22. Kortus, J., Mazin, I. I., Belashchenko, K. D., Antropov, V. P. & Boyer, L. L. Superconductivity of metallic boron in MgB_2 . *Phys. Rev. Lett.* **86**, 4656–4659 (2001).
23. Fedorov, A. V. *et al.* Temperature dependent photoemission studies of optimally doped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$. *Phys. Rev. Lett.* **82**, 2179–2182 (1999).
24. Ding, H. *et al.* Momentum dependence of the superconducting gap in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$. *Phys. Rev. Lett.* **74**, 2784–2787 (1995).
25. Reinert, F. *et al.* Observation of a BCS spectral function in a conventional superconductor by photoelectron spectroscopy. *Phys. Rev. Lett.* **87**, 047001 (2001).
26. Machida, Y. *et al.* Ambient-pressure synthesis of single-crystal MgB_2 and their superconducting anisotropy. *Phys. Rev. B* **67**, 094507 (2003).

Acknowledgements We thank J. Akimitsu for discussions. We also thank H. Höchst for experimental assistance. This work was supported by grants from the MEXT of Japan, JSPS and US NSF. S.S. thanks JSPS for financial support. The Synchrotron Radiation Center is supported by US NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.T. (t.takahashi@arps.phys.tohoku.ac.jp).

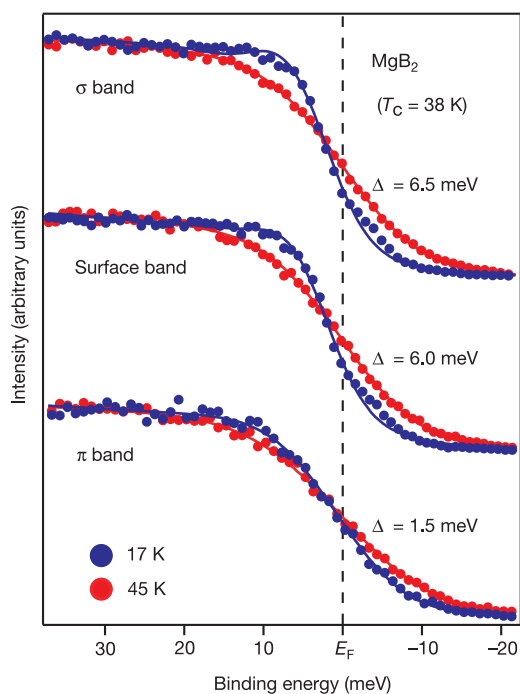


Figure 2 Temperature dependence of ARPES spectra near E_F of MgB_2 . Spectra are measured at three points A, B and C in Fig. 1b, which correspond to the π , σ and surface bands, respectively. Measurements were done at 17 K (blue circles) and 45 K (red circles). Spectra are normalized by the area under the curve from 50 meV above E_F to 200 meV below E_F . Blue and red lines on the spectra show the results of numerical fitting at 17 K and 45 K, respectively. The reproducibility of experimental data has been confirmed by measuring the spectra with the cycling temperature of the sample across T_C .

Extreme crustal oxygen isotope signatures preserved in coesite in diamond

Daniel J. Schulze^{*}, Ben Harte[†], John W. Valley[‡], James M. Brenan[§] & Dominic M. De R. Channer^{||}

^{*} Department of Geology, University of Toronto, Erindale College, Mississauga, Ontario, Canada L5L 1C6

[†] Department of Geology and Geophysics, University of Edinburgh, King's Buildings, Edinburgh EH9 3JW, UK

[‡] Department of Geology and Geophysics, University of Wisconsin, Madison, Wisconsin 53706, USA

[§] Department of Geology, University of Toronto, Toronto, Ontario, Canada M5S 3B1
^{||} Guaniamo Mining Company, Centro Mohedano, La Castellana, Caracas 1060, Venezuela

The anomalously high and low oxygen isotope values observed in eclogite xenoliths from the upper mantle beneath cratons have been interpreted as indicating that the parent rock of the eclogites experienced alteration on the ancient sea floor¹. Recognition of this genetic lineage has provided the foundation for a model of the evolution of the continents whereby imbricated slabs of oceanic lithosphere underpin and promote stabilization of early cratons². Early crustal growth is thought to have been enhanced by the addition of slab-derived magmas, leaving an eclogite residuum in the upper mantle beneath the cratons³. But the oxygen isotope anomalies observed in eclogite xenoliths are small relative to those in altered ocean-floor basalt and intermediate-stage subduction-zone eclogites, and this has hindered acceptance of the hypothesis that the eclogite xenoliths represent subducted and metamorphosed ocean-floor basalts. We present here the oxygen isotope composition of eclogitic mineral inclusions, analysed *in situ* in diamonds using an ion microprobe/secondary ion mass spectrometer. The oxygen isotope values of coesite (a polymorph of SiO₂) inclusions are substantially higher than previously reported for xenoliths from the subcratonic mantle, but are typical of subduction-zone meta-basalts, and accordingly provide strong support for the link between altered ocean-floor basalts and mantle eclogite xenoliths.

Mantle peridotites and fresh basaltic rocks have oxygen isotope ratios ($\delta^{18}\text{O}_{\text{VSMOW}}$) primarily in the range +5.5‰ to +5.9‰ (refs 4, 5), tightly constraining the isotopic composition of common mantle oxygen. Mantle-derived eclogite xenoliths clearly depart from this range, however, as illustrated by the variation of the $\delta^{18}\text{O}$ values of their garnets (approximately in the range +2‰ to +9‰) in Fig. 1. Such anomalous oxygen isotope signatures were initially interpreted as the products of fractional crystallization of mantle magmas⁶, an origin also considered by subsequent workers⁷, although oxygen isotope fractionation effects are now known to be small under mantle conditions⁸. Many more recent studies of the isotope and trace-element geochemistry of mantle eclogite xenoliths (see, for example, refs 1, 9, 10) and certain other mantle pyroxenites¹¹ have interpreted the anomalous oxygen isotope ratios as evidence that these rocks are derived from subducted altered ocean-floor basalts and gabbros.

The changes in oxygen isotope ratios and other geochemical characteristics of oceanic lithosphere that occur through exchange with sea water and hydrothermal fluids near ocean ridges have been well-documented in studies of ocean-floor rocks^{12,13} and on-land obducted ophiolite equivalents^{14,15}. Prograde metamorphic dehydration reactions do not cause large shifts in oxygen isotope ratios¹⁶, and equivalent basic rocks that have been metamorphosed at high pressures and low temperatures during subduction, but then

returned to the surface of the Earth, retain similarly anomalous oxygen isotope signatures as evidence of their oceanic heritage^{17,18}. Although these features clearly support the possibility of ocean-floor protoliths for mantle eclogite xenoliths, there has not been a close match between the upper limit of oxygen isotope ratios of mantle eclogite xenoliths and those of altered oceanic crust.

The $\delta^{18}\text{O}$ values of mantle eclogite xenoliths rarely exceed +8.0‰ (refs 6, 7, 10) and do not approach those of some of the suggested altered basaltic protoliths, in which values of $\delta^{18}\text{O} > +10.0\text{‰}$ are common (Fig. 1). The fact that the anomalous $\delta^{18}\text{O}$ values of the eclogite xenoliths are considerably lower than those of such oceanic meta-basalts has contributed to the reluctance of some workers to accept that the xenolith eclogites and their diamonds represent subducted meta-basalts tectonically emplaced beneath the cratons^{19,20}. The question therefore arises whether the anomalous oxygen isotope ratios of the eclogites are due to processes other than subduction of altered ocean-floor basalt⁷. Alternatively, it may be that, although some mantle eclogites once had extremely elevated $\delta^{18}\text{O}$ values (for example, +10 to +20‰), these signatures have been erased or modified since emplacement through processes in the mantle such as partial melting, diffusion and fluid transfer, and the current moderately elevated $\delta^{18}\text{O}$ values are remnants of formerly much higher values.

One type of sample that can help to clarify the oxygen isotope history of mantle eclogites is the suite of minerals that occur as inclusions in diamond. The study of diamond inclusion minerals has provided valuable data bearing on various aspects of diamond geology, including information on the chemical environment,

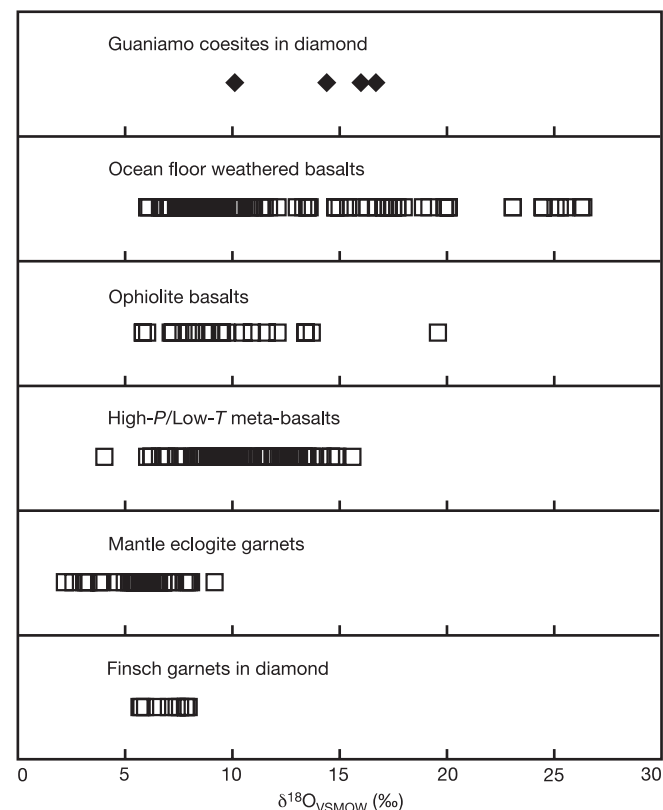


Figure 1 Oxygen isotope compositions of diamond inclusion minerals and of various basaltic rocks. Data for coesite inclusions in Guaniamo diamonds (this work) are compared with those of whole rocks (WRs) from ocean floor weathered basalts^{12,13}, WR basalts from ophiolites^{14,15}, meta-basalts from high-pressure, low-temperature, subducted metamorphic terranes (WR data from ref. 18, and garnets from ref. 17), garnets from mantle eclogite xenoliths^{6,7} and Finsch eclogite-suite garnets in diamonds²². Although data from a limited number of studies are shown here for each rock type, those chosen illustrate the range of oxygen isotope values typical of each type.

pressure, temperature and age of diamond formation²¹. Minerals encapsulated in impermeable and unreactive diamond commonly preserve a pristine record of ambient conditions during diamond formation, in contrast to the 'exposed' minerals in rocks, which have had the opportunity for exchange with surrounding material in the open system of the upper mantle.

The fact that most minerals included in diamonds are very small (typically tens to a few hundred micrometres) has limited the variety of techniques that can be applied to diamond-inclusion investigations, and until recently²² precluded oxygen isotope studies. The small beam diameter and high sensitivity of an ion microprobe, however, allows data to be collected from samples as tiny as minerals included in diamonds, and permits analysis of grains that are orders of magnitude smaller (nanograms) than possible by other methods ($\geq 50 \mu\text{g}$; ref 22). Using an ion microprobe and secondary-ion mass spectrometry (SIMS), we have now determined the oxygen isotope composition of mineral inclusions in eclogite-suite diamonds from the Guaniamo region of Venezuela, located on the Guyana shield. We chose to examine this suite because a number of features mark it as unusual. These include the facts that diamonds from this locality overwhelmingly belong to the eclogite suite, that coesite is unusually abundant as a diamond inclusion mineral, and that many of the diamonds have unusually low carbon isotope values^{23,24}. The low $\delta^{13}\text{C}$ values of the Guaniamo diamonds (as low as -28.7‰ PDB, with most diamonds in the range -10‰ to -20‰) suggest that a significant crustal component might have contributed to diamond genesis^{23,24}, and the abundance of coesite in the eclogite suite is consistent with little or no partial melting of the oceanic slab following subduction. These characteristics suggested that anomalous oxygen isotope signatures might be preserved in the Guaniamo diamond inclusion minerals.

In a small suite of Guaniamo diamonds polished to expose mineral inclusions, we identified three inclusions of SiO_2 composition in two diamonds by electron microprobe methods at the University of Toronto—their identity as coesite was established by electron backscatter diffraction at the University of Edinburgh. Inclusions in diamond 13-127-16 consist of two coesites, two omphacites and a kyanite. Diamond 13-127-X contains a single grain of coesite.

The oxygen isotope compositions of the three coesite grains were measured using a Cameca ims-4f ion microprobe at the University of Edinburgh, using a primary beam of $^{133}\text{Cs}^+$ ions defocused to a spot $\sim 30 \mu\text{m}$ in diameter and with an impact potential of 14,150 eV. Secondary ions were measured with high energy off-set. Descriptions of the oxygen isotope analytical techniques used can be found in refs 25 and 26. The coesite measurements were standardized against the Bogala quartz standard ($\delta^{18}\text{O} = +12.34\text{‰}$; ref 27). Nineteen analyses on different spots of Bogala quartz acquired in the same session and bracketing the coesite measurements yield a precision of $\pm 1.0\text{‰}$ (all errors are given as one standard deviation). The coesite data are given in Table 1. Three analyses of two coesites in diamond 13-127-16 ($\delta^{18}\text{O} = +15.0\text{‰}$ and $+16.4\text{‰}$ in one grain, and $+16.9\text{‰}$ in the second grain) yield an average of $+16.1 \pm 1.0\text{‰}$. A single analysis of the coesite in diamond 13-127-X has $\delta^{18}\text{O} = +10.2\text{‰}$. The raw SIMS data for the standards and unknowns are available as Supplementary Information.

Accurate SIMS analysis of $\delta^{18}\text{O}$ requires the use of standards similar in composition to that of the unknown²⁵. We used a quartz standard to analyse coesite, as both minerals are polymorphs of

silica. To verify that quartz is in fact a suitable standard for SIMS analysis of coesite, we have analysed synthetic coesite as a secondary standard. The synthetic coesite was prepared at the University of Toronto from NCSU quartz standard ($\delta^{18}\text{O} = +11.60\text{‰}$; ref 28). The finely powdered quartz sample was encapsulated in graphite, and converted to coesite by annealing at $1,000^\circ\text{C}$ and 3.5 GPa for 24 h using a piston-cylinder apparatus. Sample conversion was confirmed by X-ray diffraction of a portion of the run product.

The oxygen isotope composition ($\delta^{18}\text{O}$) of the synthetic coesite was determined by laser-fluorination techniques at the University of Wisconsin to be $11.63 \pm 0.33\text{‰}$, identical to that of the quartz starting material. Using the Bogala quartz standard, as for the coesite inclusions in diamond, the synthetic coesite was analysed by SIMS at the University of Edinburgh. Before SIMS analysis, the structural identity of the material at all the ion probe points was confirmed as coesite by electron backscatter diffraction. Referenced to 12 analyses of the Bogala quartz ($\delta^{18}\text{O} = +12.34 \pm 1.04\text{‰}$) bracketing the synthetic coesite analyses, nine analyses of the synthetic coesite yielded a value of $\delta^{18}\text{O} = +11.04 \pm 1.42\text{‰}$, identical within error to the laser-fluorination determinations of the synthetic coesite and the NCSU quartz. This confirms that there is no measurable instrumental mass fractionation between α -quartz and coesite, and that the Bogala quartz standard is suitable to use for SIMS analysis of the oxygen isotope composition of coesite.

Although equilibrium fractionation of $^{18}\text{O}/^{16}\text{O}$ between coexisting phases hinders direct comparison of the $\delta^{18}\text{O}$ values of coesite with the larger database of $\delta^{18}\text{O}$ values of eclogitic garnets, the extent of such fractionation will be small at mantle temperatures. Estimated equilibration temperatures for Guaniamo eclogitic diamonds are in the range $950\text{--}1,450^\circ\text{C}$ (refs 23, 24). For garnets with $\text{Ca:Mg:Fe} = 20:40:40$ (similar to the approximate average composition of garnets in Guaniamo eclogite-suite diamonds), at $1,100^\circ\text{C}$, $\delta^{18}\text{O}$ of quartz should be approximately 1.5‰ higher than coexisting garnet, and at higher temperatures the difference is smaller²⁹. Thus, $\delta^{18}\text{O}$ of (hypothetical) garnet coexisting in equilibrium with coesite ($\delta^{18}\text{O} = +16.1\text{‰}$) in diamond 13-127-16 would be near 14.6‰ (if $\delta^{18}\text{O}_{\text{quartz}} - \delta^{18}\text{O}_{\text{coesite}} \approx 0$).

Although we do not know with certainty that the coesites with highly anomalous oxygen isotope ratios formed in equilibrium with garnet, there is no reason to suspect otherwise. All published laser-fluorination oxygen isotope data for mantle eclogite xenoliths demonstrate equilibrium $\delta^{18}\text{O}$ clinopyroxene-garnet fractionation¹⁰. Laser-fluorination data for all primary phases in a coesite-kyanite eclogite xenolith also show equilibrium $\delta^{18}\text{O}$ fractionation³⁰. In addition, the two coesite grains in diamond 13-127-16 have identical $\delta^{18}\text{O}$ (within analytical error), though they are not touching, suggesting that their $\delta^{18}\text{O}$ represents the equilibrium value of coesite of the host eclogite during diamond formation. In a similar situation, the lack of sulphur isotope variability among sulphide inclusions within single diamonds from Orapa (Botswana) was interpreted to represent the equilibrium sulphur isotope composition of the immediate region in which the diamond was formed³¹. The substantial $\delta^{18}\text{O}$ difference between coesites from the two diamonds in this investigation demonstrates disequilibrium on a larger scale, however, which is typical of all suites of mantle eclogite xenoliths (see, for example, refs 6, 10).

Lowry *et al.*²² suggested the possibility of disequilibrium $\delta^{18}\text{O}$ fractionation associated with cooling or local fluid buffering during diamond formation as the cause of small ($<0.5\text{‰}$) $\delta^{18}\text{O}$ anomalies in some diamondiferous samples. Such effects were only observed in the peridotite diamond suite (garnet-olivine apparent disequilibrium), however, where the presence of Cr was considered to influence $\delta^{18}\text{O}$ fractionation. These effects have not been observed in the eclogite-diamond association¹⁰, so they are not applicable to the Guaniamo eclogite diamond population and are unlikely to be the cause of the anomalous coesite oxygen isotope ratios.

Few oxygen isotope data are available for mineral inclusions in

Table 1 Oxygen isotope values of coesite inclusions in Guaniamo diamonds

Diamond	Inclusion	$\delta^{18}\text{O}_{\text{VSMOW}} (\text{‰})$
13-127-16	3	15.0
13-127-16	3 (repeat)	16.4
13-127-16	4	16.9
13-127-X	1	10.2

diamond for comparison, and only one study has dealt with those of the eclogite suite. Ten eclogitic garnets included in Finsch diamonds have moderately anomalous $\delta^{18}\text{O}$ values ($\delta^{18}\text{O} = +5.7$ to $+8.0\text{‰}$) relative to garnet in peridotite xenoliths ($\delta^{18}\text{O} = +5.35 \pm 0.18\text{‰}$), but these data are within the range of published oxygen isotope values for garnet in mantle eclogite xenoliths²². That is, the Finsch eclogite-suite inclusions are somewhat anomalous relative to typical mantle values or 'pristine' basalts, but not unusual compared to mantle eclogite xenoliths with moderately elevated oxygen isotope values ($\delta^{18}\text{O} < +9.2\text{‰}$). Lowry *et al.*²² concluded that, although fluids associated with diamond formation in eclogites are significantly ^{18}O enriched ($> +7\text{‰}$), "the diamondiferous xenoliths have not undergone any gross exchange of $\delta^{18}\text{O}$ since the time of diamond growth and inclusion entrapment".

We have confirmed that eclogite diamond formation is associated with elevated $\delta^{18}\text{O}$ values. We have also documented that $\delta^{18}\text{O}$ values do in fact exist in members of the subcratonic eclogite suite that are equivalent to the extreme $\delta^{18}\text{O}$ values found in low-temperature altered oceanic basalts and their ophiolitic equivalents (Fig. 1). The fact that the coesite oxygen isotope ratios are outside the range of $\delta^{18}\text{O}$ known from eclogite xenoliths, however, suggests that following diamond formation, 'exposed' minerals have undergone oxygen isotope exchange with the surrounding upper mantle, and only the encapsulated minerals preserve the $\delta^{18}\text{O}$ of diamond formation.

The prospect that disequilibrium or small, but poorly known, fractionation effects can cause small shifts from expected equilibrium $\delta^{18}\text{O}$ fractionation in upper-mantle materials continues to leave open the possibility that the moderately anomalous oxygen isotope ratios of diamondiferous (and other) eclogites documented to date ($+6$ to $+9\text{‰}$) might be the result of processes other than subduction and prograde metamorphism of hydrothermally-altered oceanic basalt^{19,20}. We consider this possibility to be unlikely, but even if it is correct, such shifts would be small^{8,22}. Because of the magnitude of the coesite $\delta^{18}\text{O}$ anomalies, none of these alternative arguments apply to the extreme oxygen isotope ratios that we have documented, and no other mechanisms have been proposed to account for them. The Guaniamo coesite inclusions in diamond clearly represent the 'missing link' between altered ocean-floor basalts and mantle eclogite xenoliths, and they present compelling evidence for the subducted nature of the protoliths of this diamond eclogite suite. □

Received 2 October 2002; accepted 25 March 2003; doi:10.1038/nature01615.

- Jagoutz, E., Dawson, J. B., Hoernes, S., Spettel, B. & Wanke, H. Anorthositic oceanic crust in the Archean. *Lunar Planet. Sci.* **15**, 395–396 (1984).
- Helmstaedt, H. & Schulze, D. J. in *Kimberlites and Related Rocks Vol. 1, Their Composition, Origin and Emplacement* (ed. Ross, J.) 358–368 (Blackwell, Carlton, Australia, 1989).
- Rollinson, H. Eclogitic xenoliths in west Africa kimberlites as residues from Archean granitoid crust formation. *Nature* **389**, 173–176 (1997).
- Ito, E., White, W. M. & Gopel, C. The O, Sr, Nd, and Pb isotope geochemistry of MORB. *Chem. Geol.* **62**, 157–176 (1987).
- Mattey, D. P., Lowry, D., Macpherson, C. G. & Chazot, G. Oxygen isotope composition of mantle minerals by laser fluorination analysis: homogeneity in peridotites, heterogeneity in eclogites. *Mineral. Mag.* **58**, 573–574 (1994).
- Garlick, G. D., MacGregor, I. D. & Vogel, D. E. Oxygen isotope ratios in eclogites from kimberlites. *Science* **172**, 1025–1027 (1971).
- Deines, P., Harris, J. W., Robinson, D. N., Gurney, J. J. & Shee, S. R. Carbon and oxygen isotope variations in diamond and graphite eclogites from Orapa, Botswana, and the nitrogen content of their diamonds. *Geochim. Cosmochim. Acta* **55**, 515–524 (1991).
- Clayton, R. N., Goldsmith, J. R., Karel, K. J., Mayeda, T. K. & Newton, R. P. Limits on the effect of pressure in isotopic fractionation. *Geochim. Cosmochim. Acta* **39**, 1197–1201 (1975).
- MacGregor, I. D. & Manton, W. I. Roberts Victor eclogites: Ancient oceanic crust. *J. Geophys. Res.* **91**, 14063–14079 (1986).
- Jacob, D., Jagoutz, E., Lowry, D., Mattey, D. & Kudrjatzseva, G. Diamondiferous eclogites from Siberia: Remnants of Archean oceanic crust. *Geochim. Cosmochim. Acta* **58**, 5191–5207 (1994).
- Pearson, D. G., Davies, G. R., Nixon, P. H., Greenwood, P. G. & Mattey, D. P. Oxygen isotope evidence for the origin of pyroxenites in the Beni Bousera peridotite massif, North Morocco: derivation from subducted oceanic lithosphere. *Earth Planet. Sci. Lett.* **102**, 289–301 (1991).
- Muehlenbachs, K. The alteration and aging of the basaltic layer of the sea floor: oxygen isotope evidence from DSDP/IPOD legs 51, 52, and 53. *Init. Rep. DSDP LI-LIIL*, 1159–1167 (1980).
- Staudigel, H., Muehlenbachs, K., Richardson, S. H. & Hart, S. R. Agents of low temperature ocean crust alteration. *Contrib. Mineral. Petrol.* **77**, 150–157 (1981).
- Cocker, J. D., Griffin, B. J. & Muehlenbachs, K. Oxygen and carbon isotope evidence for seawater-

- hydrothermal alteration of the Macquarie Island ophiolite. *Earth Planet. Sci. Lett.* **61**, 112–122 (1982).
- Schiffman, P., Williams, A. E. & Evarts, R. C. Oxygen isotope evidence for submarine hydrothermal alteration of the Del Puerto ophiolite, California. *Earth Planet. Sci. Lett.* **70**, 207–220 (1984).
- Valley, J. W. in *Stable Isotopes in High Temperature Geological Processes* (eds Valley, J. W., Taylor, H. P. & O'Neill, J. R.) 445–489 (Reviews in Mineralogy No. 16, Mineralogical Society of America, Washington DC, 1986).
- Leech, M. L. & Ernst, W. G. Petrotectonic evolution of the high- to ultrahigh-pressure Maksyutov Complex, Karayanova area, south Ural Mountains: structural and oxygen isotope constraints. *Lithos* **52**, 235–252 (2000).
- Miller, J. A., Cartwright, I., Buick, I. S. & Bairncoat, A. C. An O-isotope profile through the HP-LT Corsican ophiolite, France and its implications for fluid flow during subduction. *Chem. Geol.* **178**, 43–69 (2001).
- Haggerty, S. E. A diamond trilogy: superplumes, supercontinents, and supernovae. *Science* **285**, 851–860 (1999).
- Cartigny, P., Harris, J. W. & Javoy, M. in *Proc. VIIth Int. Kimberlite Conf.* Vol. 1 (eds Gurney, J. J., Gurney, J. L., Pascoe, M. D. & Richardson, S. H.) 117–124 (Red Roof Design, Cape Town, 1999).
- Gurney, J. J. in *Kimberlites and Related Rocks Vol. 2, Their Mantle/Crust Setting, Diamonds and Diamond Exploration* (ed. Ross, J.) 935–965 (Blackwell, Carlton, Australia, 1989).
- Lowry, D., Mattey, D. P. & Harris, J. W. Oxygen isotope composition of syngenetic inclusions in diamond from the Finsch Mine, RSA. *Geochim. Cosmochim. Acta* **63**, 1825–1836 (1999).
- Sobolev, N. V., Efimova, E. S., Channer, D. M. de R., Anderson, P. F. N. & Barron, K. M. Unusual upper mantle beneath Guaniamo, Guyana Shield, Venezuela: evidence from diamond inclusions. *Geology* **26**, 971–974 (1998).
- Kaminsky, F. V., Zakharchenko, O. D., Griffin, W. L., Channer, D. M. de R. & Khachatryan-Blinova, G. K. Diamond from the Guaniamo area, Venezuela. *Can. Mineral.* **38**, 1347–1370 (2000).
- Eiler, J. M., Graham, C. & Valley, J. W. SIMS analysis of oxygen isotopes: matrix effects in complex minerals and glasses. *Chem. Geol.* **138**, 221–244 (1997).
- Valley, J. W., Graham, C. M., Harte, B., Eiler, J. M. & Kinny, P. D. in *Applications of Microanalytical Techniques to Understanding Mineralizing Processes* (eds McKibben, M. A., Shanks, W. C. III & Ridley, W. I.) 73–98 (SEG Reviews in Economic Geology Vol. 7, Society of Economic Geologists, 1998).
- Valley, J. W. & Graham, C. Ion microprobe analysis of oxygen isotope ratios in quartz from Skye granite: healed micro-cracks, fluid flow, and hydrothermal exchange. *Contrib. Mineral. Petrol.* **124**, 225–234 (1996).
- Spicuzza, M. J., Valley, J. W., Kohn, M. J., Gierard, J. P. & Fouillac, A. M. The rapid heating, defocused beam technique: A CO₂-laser based method for highly precise and accurate determination of $\delta^{18}\text{O}$ values of quartz. *Chem. Geol.* **144**, 195–203 (1998).
- Valley, J. W., Bindeman, I. N. & Peck, W. H. Empirical calibration of oxygen isotope fractionation in zircon. *Geochim. Cosmochim. Acta* (in the press).
- Sharp, Z. D., Essene, E. J. & Smyth, J. R. Ultra-high temperatures from oxygen isotope thermometry of a coesite-sanidine grosspyrite. *Contrib. Mineral. Petrol.* **112**, 358–370 (1992).
- Farquhar, J. *et al.* Mass-independent sulfur of inclusions in diamond and sulfur recycling on early Earth. *Science* **298**, 2369–2372 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Craven for invaluable technical assistance with the ion microprobe. We also thank R. Cooper for access to the diamonds, N. Cayzer for assistance with general scanning electron microscopy work and electron backscatter diffraction determinations on the coesite, M. Spicuzza for assistance in laser fluorination and mass spectrometry, B. Schumacher and associates for polishing the diamonds, A. Dias for help with the figure, H. Halls for discussion, and G. Ernst for comments and suggestions on the manuscript. D.J.S. and J.M.B. are supported by NSERC, and J.W.V. by NSF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.S. (dschulze@credit.erin.utoronto.ca).

Migration of a Late Cretaceous fish

Scott J. Carpenter*, J. Mark Erickson† & F. D. Holland Jr‡

* Paul H. Nelson Stable Isotope Laboratory, Department of Geoscience and Center for Global and Regional Environmental Research, University of Iowa, Iowa City, Iowa 52242-1379, USA

† Geology Department, St Lawrence University, Canton, New York 13617, USA

‡ Department of Geology and Geological Engineering, University of North Dakota, Grand Forks, North Dakota 58202, USA

Late Cretaceous sediments from the Western Interior of North America yield exceptionally well preserved fossils^{1,2} that serve as proxies for the rapidly changing climate preceding the Cretaceous/Tertiary boundary (about 67–65 Myr ago)^{3,4}. Here we reconstruct the ontogenetic history of a Maastrichtian-age fish,

*Vorhisia vulpes*⁵, by using the carbon, oxygen and strontium isotope ratios of four aragonite otoliths collected from the Fox Hills Formation of South Dakota. Individuals of *V. vulpes* spawned in brackish water (about 70–80% seawater) and during their first year migrated to open marine waters of the Western Interior Seaway, where they remained for 3 years before returning to the estuary, presumably to spawn and die. The mean $\delta^{18}\text{O}$ from the marine growth phase of *V. vulpes* yields a seawater temperature of 18 °C, which is consistent with leaf physiognomy and general-circulation-model temperature estimates for the Western Interior during the latest Maastrichtian^{4,6,7}.

Modern fish otoliths have proved to be useful proxies for ambient water conditions^{8,9}. Biogeochemists have used stable-isotope time series from otoliths to reconstruct detailed ontogenetic histories^{10,11} and to characterize ambient life conditions and depositional environments^{12,13}. The isotopic analysis of ancient otoliths permits the detailed characterization of ontogenetic histories when skeletal remains are missing and classification is in question.

We have analysed four Late Cretaceous (Maastrichtian) utricular otoliths (*Vorhisia vulpes*) collected from a tidal-channel lag deposit in the Iron Lightning Member (Colgate Lithofacies) of the Fox Hills Formation of South Dakota (Fig. 1). The Iron Lightning Member was deposited in coastal estuaries that developed behind a series of barrier islands along the margin of an extensive delta platform^{14–16} and is overlain there by fluvial deposits of the Hell Creek Formation. These channel deposits represent the distal portion of the Hell Creek fluvial distributary system, where it flowed into and mixed with the Fox Hills Sea. Isotope analyses were conducted on these otoliths to understand the life history of these fish and to characterize the riverine, estuarine and marine environments of the type area of the Fox Hills Formation in the Missouri Valley Region of South and North Dakota during the latest Cretaceous¹⁷.

A discussion of the taxonomic assignment of the otoliths is appropriate because they have been left in open nomenclature by recent revisers¹⁸. These utricular otoliths were originally described from the Fox Hills Formation of South Dakota by Frizzell⁵ as *V. vulpes*, and have been cited as such in the regional literature. On the basis of presumably conspecific material from the Severn Formation in Maryland, they have recently been assigned¹⁸ to a form genus 'Ariidarum' without further description or evidentiary discussion. We choose to employ the original appellation⁵ for three

reasons: the present work is not intended to influence classificatory decisions regarding these specimens; we are using material from the type area of the species *V. vulpes* and clearly conspecific with it; and the name is well known regionally.

Carpenter *et al.*¹ described the exceptional preservation of molluscan aragonite contained within carbonate cemented concretions from the Timber Lake Member of the Fox Hills Formation and characterized $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ values of those marine molluscs. Otoliths analysed here, composed of fibrous aragonite, are equally well preserved, having undergone no diagenetic alteration, cementation or dissolution. One specimen (SLU FR478) is abraded, probably by subaqueous transport. Concentric growth bands are visible on all specimens (Fig. 2). Ultraviolet fluorescence petrography reveals an overall high organic matter (otolin?) content within each otolith and a concentration along growth bands (Supplementary Information).

Oxygen isotope time series from four specimens of *V. vulpes* (Fig. 2), collected from the same bedding-plane surface, yield similar patterns of oxygen isotope variation that result from migration during the ontogeny of each fish. Although these specimens do not yield geochemical data from synchronous growth, they represent specimens that probably lived within tens or hundreds of years of each other. We assume that these data are representative of the habitats and ontogenetic history of *V. vulpes*. The rapid growth rates of *V. vulpes* lapilli (1.1–1.25 cm maximum length; 0.4–0.5 g total mass) for a maximum age of 4 years, would place them among the highest of modern teleosts (Supplementary Information).

Aragonite precipitated in the kernel (first ~1.5 mm of growth) of all otoliths analysed has relatively low $\delta^{18}\text{O}$ values (from –6.2‰ to –3.4‰), typical of growth in an estuarine habitat. Between 2 and 4 mm of growth, $\delta^{18}\text{O}$ values increase abruptly to a mean of –1.1‰, typical of otolith growth in a marine environment. The marine phase of otolith growth is characterized by a sinusoidal variation in $\delta^{18}\text{O}$ values about this mean (Fig. 2). We interpret these sinusoidal patterns as seasonal temperature variation in the shallow marine waters of the western margin of the Fox Hills Sea, and conclude that *V. vulpes* lived in open marine water for about 3 years. $^{87}\text{Sr}/^{86}\text{Sr}$ ratios of 0.70778 to 0.70779 for the marine portions of *V. vulpes* otoliths are consistent with those of other marine molluscs from the Western Interior Seaway and with an age of between 67 and 65.5 Myr ago (Supplementary Information)².

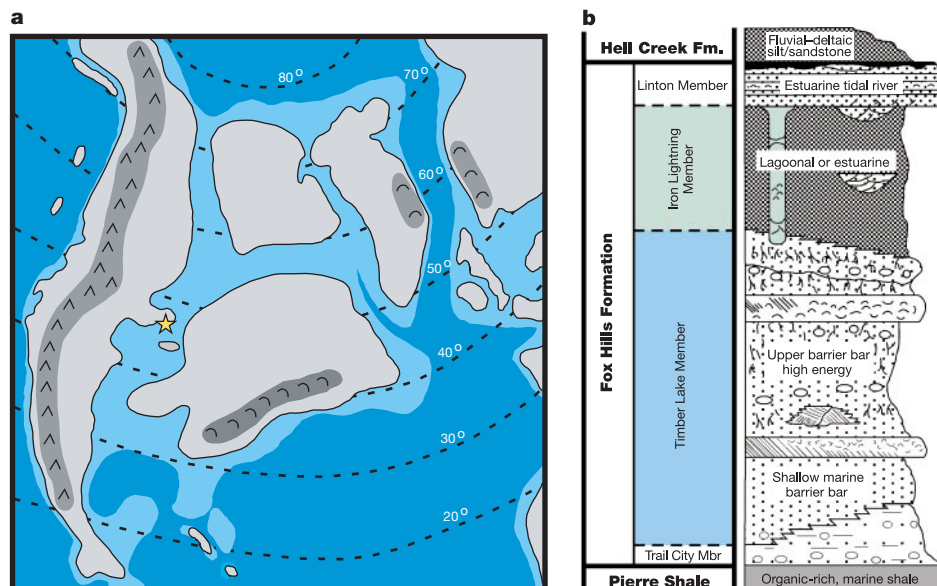


Figure 1 Palaeogeography and stratigraphy of the Fox Hills Formation of North and South Dakota. **a**, Palaeogeographic map after ref. 16. The palaeolatitude of this location is 46° N. **b**, Schematic stratigraphic column of the Pierre–Fox Hills–Hell Creek Formations of

North and South Dakota. Modified from ref. 1. *V. vulpes* specimens were collected from the estuarine tidal channel (green) in the Iron Lightning Member. The marine Timber Lake Member is shown in blue.

Using the fractionation factor of Patterson *et al.*⁸ for modern fish otoliths and a seawater $\delta^{18}\text{O}$ of -1‰ (standard mean ocean water; SMOW) (for an ice-free world¹⁹), the temperature range for all four otoliths would be $14\text{--}25\text{ °C}$. An otolith $\delta^{18}\text{O}$ of -1.1‰ yields a temperature of 18 °C (Fig. 2). These temperature estimates are significantly lower than those calculated from the Thorrold *et al.*⁹ relation (otolith $\delta^{18}\text{O}$ of -1.1‰ yields a temperature of 22 °C and an annual temperature range of $18\text{--}30\text{ °C}$). As temperatures calculated from the Thorrold *et al.*⁹ relation seem high for this palaeo-latitude, we have used the fractionation factor of Patterson *et al.*⁸ (Fig. 2). Although high relative to modern mid-latitude temperatures, a seawater temperature of 18 °C is consistent with temperature estimates for 46 °N during the Late Cretaceous³. This temperature is several degrees lower than general circulation model results for this latitude in the mid-Cretaceous period (atmospheric CO_2 levels fourfold modern levels)⁷ but is consistent with leaf physiognomy data projecting mean annual temperatures of $15\text{--}22\text{ °C}$ for the Western Interior during the late Maastrichtian^{4,6}.

The death of each fish occurred without precipitation of a final spring–summer growth increment. We therefore conclude that the return of *V. vulpes* to estuarine waters (presumably to spawn) occurred in the autumn. As these otoliths were discovered in known estuarine beds^{14–16} and because isotope ratios of the juvenile phase of otolith growth indicate precipitation in an estuary, we infer that death occurred in their natal waters. Because the return of *V. vulpes* to estuarine waters was not recorded in the $\delta^{18}\text{O}$ values of

otolith outer margins, we suggest that there might have been no otolith growth during the physiologically demanding (and short-lived) spawning interval.

On the assumption that *V. vulpes* spawned in the fall, the early record of otolith growth occurred during the following spring. The time associated with the increase in otolith $\delta^{18}\text{O}$ values during the estuarine–marine transition is difficult to determine because the change in the isotopic composition of ambient waters may obscure seasonal temperature variation. However, as transitional $\delta^{18}\text{O}$ values adjoin the marine sinusoidal pattern during the summer temperature maximum ($\delta^{18}\text{O}$ value minimum; Fig. 2), the time that the juvenile *V. vulpes* spent in the estuary is less than 1 year. Migration to seawater during the first summer of life is common in modern chinook and sockeye salmon and requires the attainment of a size capable of withstanding salinity changes^{20,21}. High rates of otolith growth are required to record this short-lived migration. Growth curves consistent with this interpretation are presented in Supplementary Information.

$\delta^{13}\text{C}$ values from the juvenile phase of otolith growth (estuarine) are low (-7‰ to -4‰) and the latest phases of otolith growth (marine) are relatively high (more than 0‰). The marine growth phase of *V. vulpes* has $\delta^{13}\text{C}$ values that are comparable with those reported for marine bivalves¹. This overall increase in otolith $\delta^{13}\text{C}$ values during ontogeny is consistent with a transition from estuarine to marine conditions, yet marked seasonal variation superimposed on an overall trend towards higher $\delta^{13}\text{C}$ values

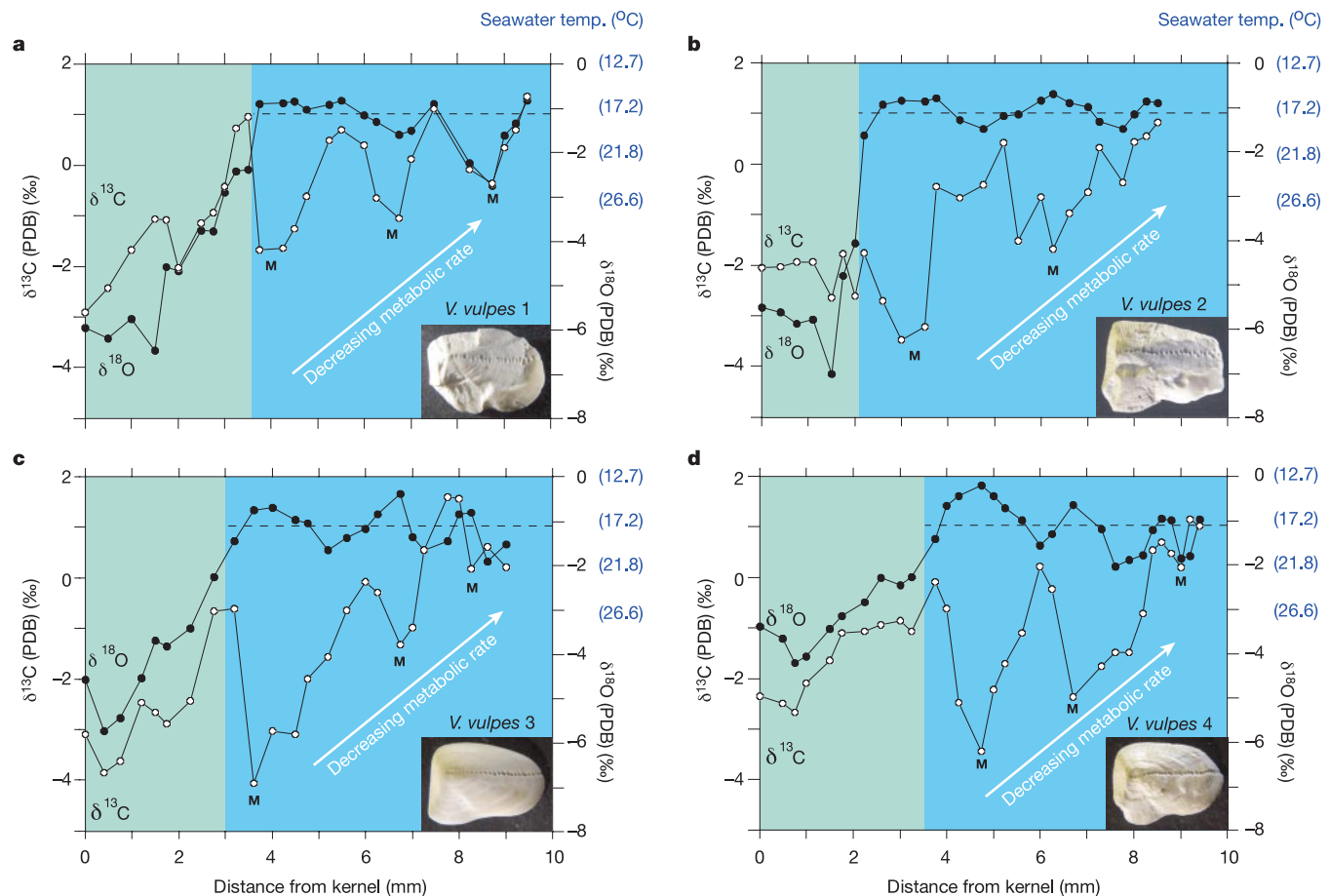


Figure 2 $\delta^{18}\text{O}$ (filled circles) and $\delta^{13}\text{C}$ (open circles) values of *Vorhisia vulpes* otolith specimens 1 (SLU FR476) (a), 2 (SLU FR477) (b), 3 (SLU FR478) (c) and 4 (SLU FR479) (d). Photographs of each specimen are found in the lower right corner of each plot. The maximum lengths of each otolith are: 1, 1.25 cm; 2, 1.10 cm; 3, 1.15 cm; 4, 1.20 cm. The green shaded area represents estuarine conditions; the blue shaded area represents marine conditions. Seawater temperature estimates using the oxygen isotope

fractionation factor of ref. 8 and a seawater $\delta^{18}\text{O}$ of -1‰ are next to the highest oxygen isotope axis labels. Temperature estimates apply to the marine $\delta^{18}\text{O}$ values in the blue shaded areas. The horizontal dashed line represents the mean $\delta^{18}\text{O}$ and temperature of the marine phase of *V. vulpes* (-1.1‰ , 17.6 °C). $\delta^{13}\text{C}$ minima indicated by M are interpreted as having been caused by annual migrations within the basin.

suggests that this environmental change does not fully explain the observed variation. Carbon-isotope time series collected from *V. vulpes* must be interpreted in terms of environment, diet, metabolic activity and migration (Fig. 2).

Dietary changes associated with a change from feeding at lower trophic levels to feeding at higher trophic levels during ontogeny might account for some of the increase in *V. vulpes* otolith $\delta^{13}\text{C}$ values with increasing age²². However, trophic level changes in $\delta^{13}\text{C}$ values are typically modest (up to 1‰ per trophic level)²² and would probably not account for all of the ontogenetic variation observed in *V. vulpes* otoliths (Fig. 2). Otoliths of juvenile chinook salmon from hatchery-raised fish have significantly higher $\delta^{13}\text{C}$ values than those of naturally spawned 'wild' juveniles. This reflects the concomitant differences in $\delta^{13}\text{C}$ values of hatchery diets between those that include a large marine component and those that include freshwater aquatic insects consumed by wild young (with lower $\delta^{13}\text{C}$ values) (W. P. Patterson, personal communication).

The overall increase in $\delta^{13}\text{C}$ values in *V. vulpes* otoliths is the result of decreasing metabolic activity^{23–25} and potentially of trophic level changes^{23,24} during ontogeny. Atlantic croaker otolith $\delta^{13}\text{C}$ values have been described⁹ that are 3–8‰ lower than ambient dissolved inorganic carbon of seawater. Ontogenetic variation similar to that observed in *V. vulpes* (Fig. 2) has been observed in the freshwater drum²⁵. These trends have been interpreted as a diminished contribution of respired CO_2 to endolymph HCO_3^- as a result of decreasing metabolic activity during ontogeny²⁵. If we assume that the timing of a maximum otolith $\delta^{13}\text{C}$ value marks sexual maturation as in Atlantic cod²³, maturation of *V. vulpes* occurred just before spawning. Although high $\delta^{13}\text{C}$ values are achieved, *V. vulpes* otoliths do not attain a $\delta^{13}\text{C}$ plateau as seen in cod and drum otoliths, probably because of the short life span of *V. vulpes* (about 4 years) and the rigors of a final spawning migration soon after sexual maturation.

Superimposed on this ontogenetic trend are periodic marked decreases in $\delta^{13}\text{C}$ values at 2–5, 5–7 and 7–9 mm (Fig. 2). With the exception of otolith no. 1, which has a poorly defined sinusoidal trend in $\delta^{18}\text{O}$ during marine growth, these $\delta^{13}\text{C}$ minima are coincident with annual temperature minima. Such decreases (1–4‰) seem too large to be caused solely by seasonal variation in primary productivity. Seasonal carbon isotope variation in North Atlantic cod otoliths²⁴ is smaller than that of freshwater drum otoliths²⁵ (1–2‰ compared with 4–5‰), presumably owing to the greater impact that primary productivity has on the $\delta^{13}\text{C}$ of ambient dissolved inorganic carbon in freshwater environments. $\delta^{13}\text{C}$ variation in *V. vulpes* (about 2–4‰; Fig. 2) might, in part, reflect the shallow nature of the Fox Hills Sea and its sensitivity to seasonal changes in productivity.

Periodic $\delta^{13}\text{C}$ minima in *V. vulpes* are interpreted as episodes of elevated metabolism or catabolism associated with migration or elevated predation stress and an increased contribution of respired CO_2 to the endolymph calcifying fluid (Fig. 2). Physiological changes associated with migration into seawater (osmoregulatory stress and/or increased metabolic activity)²⁶ are the likely cause of the first of these $\delta^{13}\text{C}$ -lowering events during ontogeny (at 2–5 mm; Fig. 2). A correlation between low temperatures and low $\delta^{13}\text{C}$ values runs counter to the positive correlation between temperature and metabolic activity, which generally lowers otolith $\delta^{13}\text{C}$ values⁹.

Decreases in otolith $\delta^{13}\text{C}$ values coincident with temperature minima suggest that *V. vulpes* migrated seasonally within the marine basin, perhaps because of changes in surface-water productivity and the diminished availability of food during the winter months (Fig. 2). Modern chinook salmon achieve migration speeds approaching 80 km d^{-1} (ref. 21). Maintenance of such migration speeds for an extended period would probably require a significant loss of body mass through catabolism. Decreases in otolith $\delta^{13}\text{C}$ values also would be expected during periods of gametogenesis and spawning²⁷. However, given that *V. vulpes* otoliths occur in

decidedly estuarine sediments, that they are of uniform size and weight and that their early life history is estuarine, it is likely that periodic $\delta^{13}\text{C}$ minima do not represent spawning migrations.

Concentrations of *V. vulpes* otoliths of a similar age class in estuarine deposits suggest that *V. vulpes* migrated back to their natal waters to spawn and then die. If spawning had occurred at a variety of ages, a range of otolith sizes would probably be found. *V. vulpes* otoliths collected from the Fox Hills Formation^{5,14} and the Severn Formation of Maryland²⁸ have a similar size distribution, suggesting that this behaviour is typical for this fish.

$\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ values of marine, estuarine and freshwater biogenic carbonates from the Fox Hills and Hell Creek Formations produce a linear mixing trend interpreted as the mixing of Fox Hills seawater and Hell Creek river water over a range of about 20‰ for $\delta^{18}\text{O}$ values and about 7‰ for $\delta^{13}\text{C}$ values (Supplementary Information)¹⁷. With the use of a two-component mixing model with a seawater $\delta^{18}\text{O}$ of –1‰ (SMOW)¹⁹ (marine molluscs and marine phases of *V. vulpes*) and a river-water $\delta^{18}\text{O}$ of –20‰ (SMOW)²⁹ (unionids and *Corbicula*), the lowest $\delta^{18}\text{O}$ values of the juvenile growth phases of the *V. vulpes* (–7.0‰, –6.5‰, –5.7‰ and –4.2‰ (PDB)) indicate that these fish initially grew in waters composed of 68–83% seawater.

The remarkable preservation of *V. vulpes* otoliths provides a detailed record of both the ontogeny of this fish and the palaeo-environmental conditions of its habitat. Extracting such information from Mesozoic-age otoliths might provide important information for phylogenetic assignments of this and other enigmatic taxa. Our data indicate that *V. vulpes* did not live in fresh water as previously suggested^{5,14}, but instead preferred shallow marine waters for much of its adult life and spawned in brackish waters. There is no evidence to suggest that *V. vulpes* lived for an extended period in the isotopically distinct freshwaters of the Hell Creek delta platform.

The absence of Late Maastrichtian ammonite biostratigraphic indices in upper Fox Hills–Hell Creek estuarine-fluvial beds enhances the value of isotope data of contained fossils for palaeoenvironmental and chemostratigraphic uses. An improved understanding of surface-water isotope ratios and temperatures (both terrestrial and marine) provides constraints for climate modellers and vertebrate palaeontologists reconstructing habitats of the Western Interior at the close of the Cretaceous. □

Methods

V. vulpes specimens 1–4 (SLU FR476–479 respectively), *Euspira subscassa* (SLU FR480) and *Corbicula* sp. (SLU FR481, FR482) are part of the St Lawrence University Faculty Research Collection collected in the 1960s from the Colgate Lithofacies, Fox Hills Formation, Iron Lightning Badlands (type locality of the Iron Lightning Member)¹⁵, Redelm NE Quadrangle, Ziebach County, South Dakota. Additional stratigraphic information is provided in the Supplementary Information.

Microsamples of aragonite (~0.02–0.05 mg) were milled from the surface of the inner faces of otoliths with a Brüsseler UP 200 controller and a UG 12 handpiece fitted with a 0.3-mm tungsten carbide dental bur with observation under a Nikon SMZU microscope. Samples were collected along transects perpendicular to visible growth bands from the posterior to the anterior of each otolith. Samples were analysed on a Finnigan MAT 252 Isotope Ratio Mass Spectrometer with a Kiel III automated carbonate device housed in the Paul H. Nelson Stable Isotope Laboratory at the University of Iowa. Carbonates were reacted with two drops of anhydrous phosphoric acid at 75 °C. Daily analyses of National Institute of Standards and Technology powdered carbonate standards (NBS-18, NBS-19 and NBS-20) and several in-house standards were conducted. Analytical precision on these standards was better than $\pm 0.1\text{‰}$ for both $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$. All results are reported in per ml (‰) relative to V-PDB.

Microsamples of aragonite were dissolved in ultrapure 3 M HNO_3 and loaded onto chromatography columns containing Sr-specific resin (EiChromM Sr Resin SPS); Sr was recovered in the aqueous eluate. Dried extracts were dissolved in 3 μl of ultrapure dilute HNO_3 and loaded onto rhenium filaments (with Ta_2O_5 in dilute H_3PO_4). Isotopic ratios were measured on a Finnigan MAT 261 thermal ionization mass spectrometer in the Radiogenic Isotope Laboratory of the University of Texas at Dallas. Each analysis typically involved 160 static multi-collection cycles of all four Sr isotopes in addition to ^{85}Rb , for monitoring ^{87}Rb . Samples were analysed over a 4-week period during which replicate measurements of standards were conducted before and after sample analyses. Standard analyses produced the following statistics: NBS-987 ($n = 55$), mean ($2\sigma_m$) = 0.710242(12), s.d. = 0.000016; EN-1 ($n = 10$), mean ($2\sigma_m$) = 0.709180(11),

s.d. = 0.000014; E&A ($n = 8$), mean ($2\sigma_m$) = 0.708008(13), s.d. = 0.000020. A polynomial least-squares fit to NBS-987 variance during the analysis period was used to drift-compensate measured ratios relative to a baseline of 0.710242. Compensations for these samples ranged from -0.000015 to 0.000009.

Received 4 December 2002; accepted 17 March 2003; doi:10.1038/nature01575.

1. Carpenter, S. J., Erickson, J. M., Lohmann, K. C. & Owen, M. R. Diagenesis of fossiliferous concretions from the Upper Cretaceous Fox Hills Formation, North Dakota. *J. Sedim. Petrol.* **58**, 706–723 (1988).
2. McArthur, J. M., Kennedy, W. J., Chen, M., Thirlwall, M. F. & Gale, A. S. Strontium isotope stratigraphy for Late Cretaceous time: direct numerical calibration of the Sr isotope curve based on the US Western Interior. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **108**, 95–119 (1994).
3. Barrera, E. & Savin, S. in *Evolution of the Cretaceous Ocean–Climate System* (eds Barrera, E. & Johnson, C. C.) 245–282 (Geol. Soc. Am. Spec. Pap. 332, 1999).
4. Wilf, P., Johnson, K. R. & Huber, B. T. Correlated terrestrial and marine evidence for global climate changes before the mass extinction at the Cretaceous–Paleogene boundary. *Proc. Natl Acad. Sci. USA* **100**, 599–604 (2003).
5. Frizzell, D. L. Otoliths of new fish (*Vorhisia vulpes*, N. Gen., N. Sp. Siluroidei?) from Upper Cretaceous of South Dakota. *Copeia* **2**, 178–181 (1965).
6. Wolfe, J. A. & Upchurch, G. R. Jr North American nonmarine climates and vegetation during the Late Cretaceous. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **61**, 33–77 (1987).
7. Barron, E. J., Fawcett, P. J., Pollard, D. & Thompson, S. Model simulations of Cretaceous climates: the role of geography and carbon dioxide. *Phil. Trans. R. Soc. Lond. B* **341**, 307–315 (1993).
8. Patterson, W. P., Smith, G. R. & Lohmann, K. C. in *Continental Climate Change from Isotopic Indicators* (eds Swart, P., Lohmann, K. C., McKenzie, J. & Savin, S.) 191–202 (Am. Geophys. Un. Monogr., 1993).
9. Thorrold, S. R., Campana, S. E., Jones, C. M. & Swart, P. K. Factors determining $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ fractionation in aragonitic otoliths of marine fish. *Geochim. Cosmochim. Acta* **61**, 2909–2919 (1997).
10. Wurster, C. M. & Patterson, W. P. Late Holocene climate change for the eastern interior United States: evidence from high-resolution $\delta^{18}\text{O}$ values of sagittal otoliths. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **170**, 81–100 (2001).
11. Ivany, L. C., Patterson, W. P. & Lohmann, K. C. Increase in seasonality across the Eocene–Oligocene boundary inferred from high-resolution microsampling of fossil otoliths, US Gulf Coastal Plain. *Nature* **407**, 887–890 (2000).
12. Andrus, C. F. T., Crowe, D. E., Sandweiss, D. H., Reitz, E. J. & Romanek, C. S. Otolith $\delta^{18}\text{O}$ record of Mid-Holocene sea surface temperatures in Peru. *Science* **295**, 1508–1511 (2002).
13. Patterson, W. P. Oldest isotopically characterized fish otoliths provide insight to Jurassic continental climate of Europe. *Geology* **27**, 199–202 (1999).
14. Waage, K. M. *The Type Fox Hills Formation, Cretaceous (Maestrichtian), South Dakota*, Part 1. *Stratigraphy and Paleoenvironments* (Peabody Museum of Natural History, Yale University, Bulletin 27, 1968).
15. Erickson, J. M. in *Proc. F. D. Holland, Jr., Geol. Symp.* (eds Erickson, J. M. & Hoganson, J. W.) 199–241 (North Dakota Geological Survey Miscellaneous Series no. 76, 1992).
16. Erickson, J. M. The Dakota Isthmus—closing the Late Cretaceous Western Interior Seaway. *North Dakota Acad. Sci. Proc.* **53**, 124–129 (1999).
17. Carpenter, S. J., Erickson, J. M. & Hoganson, J. W. Isotopic characterization of the Late Cretaceous Fox Hills – Hell Creek Estuary of North and South Dakota. *Geol. Soc. Am. Abstr. Program* 31–32 (2002).
18. Nolf, D. & Stringer, G. L. in *Mesozoic Fishes: Systematics and Paleobiology*, *Proc. Int. Meeting* (eds Arratia, G. & Viohl, G.) 433–459 (Friedrich Pfeil, München, 1996).
19. Shackleton, N. J. & Kennett, J. P. Paleotemperature history of the Cenozoic and initiation of Antarctic glaciation: Oxygen and carbon isotope analyses in Deep Sea Drilling Project Sites 277, 279, and 281. *Init. Rep. DSDP* **74**, 761–776 (1975).
20. Randall, R. G., Healy, M. C. & Dempson, J. B. in *Common Strategies of Anadromous and Catadromous Fishes* (eds Dadswell, M. J. et al.) 27–41 (American Fisheries Society, Bethesda, Maryland, 1987).
21. Healy, M. C. & Groot, C. in *Common Strategies of Anadromous and Catadromous Fishes* (eds Dadswell, M. J. et al.) 298–312 (American Fisheries Society, Bethesda, Maryland, 1987).
22. Peterson, B. J. & Fry, B. Stable isotopes in ecosystem studies. *Annu. Rev. Ecol. Syst.* **18**, 293–320 (1987).
23. Schwarcz, H. P. et al. Stable carbon isotope variations in otoliths of Atlantic cod (*Gadus morhua*). *Can. J. Aquat. Sci.* **55**, 1798–1806 (1998).
24. Weidman, C. R. & Millner, R. High-resolution stable isotope records from North Atlantic cod. *Fisheries Res.* **46**, 327–342 (2000).
25. Wurster, C. M. & Patterson, W. P. Metabolic rate of late Holocene freshwater fish: evidence from $\delta^{13}\text{C}$ values of otoliths. *Paleobiology* (in the press).
26. McCormick, S. D. & Saunders, R. L. in *Common Strategies of Anadromous and Catadromous Fishes* (eds Dadswell, M. J. et al.) 211–229 (American Fisheries Society, Bethesda, Maryland, 1987).
27. Ursin, E. in *Symp. Zool. Soc. Lond.* **44**, 63–87 (1979).
28. Huddleston, R. W. & Savoie, K. M. Teleostean otoliths from the Late Cretaceous (Maestrichtian age) Severn Formation of Maryland. *Proc. Biol. Soc. Wash.* **96**, 658–663 (1983).
29. Dettman, D. L. & Lohmann, K. C. Oxygen isotope evidence for high-altitude snow in the Laramide Rocky Mountains of North America during the Late Cretaceous and Paleogene. *Geology* **28**, 243–246 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. Ludvigson, L. Gonzalez, T. White, H. Schwarcz and W. Patterson for comments on preliminary drafts of the manuscript, and N. Miller for strontium isotope analyses of biogenic carbonates.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.J.C. (scott-j-carpenter@uiowa.edu).

Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*

D. Tian*, M. B. Traw*, J. Q. Chen†, M. Kreitman* & J. Bergelson*

* Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street, Chicago, Illinois 60637, USA

† Department of Biology, Nanjing University, 22 Han Kou Road, Nanjing 210093, People's Republic of China

Resistance genes (R-genes) act as an immune system in plants by recognizing pathogens and inducing defensive pathways. Many R-gene loci are present in plant genomes, presumably reflecting the need to maintain a large repertoire of resistance alleles. These loci also often segregate for resistance and susceptibility alleles that natural selection has maintained as polymorphisms within a species for millions of years^{1–5}. Given the obvious advantage to an individual of being disease resistant, what prevents these resistance alleles from being driven to fixation by natural selection? A cost of resistance⁶ is one potential explanation; most models require a lower fitness of resistant individuals in the absence of pathogens for long-term persistence of susceptibility alleles⁷. Here we test for the presence of a cost of resistance at the *RPM1* locus of *Arabidopsis thaliana*. Results of a field experiment comparing the fitness of isogenic strains that differ in the presence or absence of *RPM1* and its natural promoter reveal a large cost of *RPM1*, providing the first evidence that costs contribute to the maintenance of an ancient R-gene polymorphism.

RPM1 codes for a peripheral plasma membrane protein⁸ that confers the ability to recognize *Pseudomonas syringae* pathogens carrying *AvrRpm1* or *AvrB*^{9,10}. Susceptible individuals lack the entire coding region of *RPM1* (ref. 10), so there is a single susceptible allele at this locus. Both resistance and susceptibility alleles frequently occur together within natural populations and are common across the range of *A. thaliana*³. Molecular evolutionary analysis of the non-coding DNA flanking *RPM1* indicates that the resistance and susceptibility alleles have coexisted for more than nine million years³. The great age of this polymorphism indicates that the susceptibility allele might be maintained by natural selection due to a cost of resistance, because other mechanisms of coexistence can be eliminated. First, heterozygote advantage can be excluded by *A. thaliana*'s high selfing rate¹¹. Second, there is no evidence of geographic differentiation between allelic classes³. Last, because the susceptibility allele is deleted for the entire locus¹⁰, it cannot have differential adaptive values in alternative environments.

To test for a fitness cost of *RPM1* expression, we inserted a 3.84-kilobase (kb) region from Columbia containing *RPM1*, its promoter and its full terminator¹⁰ into a susceptible ecotype, Bla-2, to create four independent transgenic lines. Transgenic plants containing this fragment have been shown to respond to infection with pathogenic *P. syringae* DC3000:*AvrRpm1* by eliciting the appropriate hypersensitive response (D. Boyes and J. L. Dangl, personal communication). The *RPM1* gene fragment was placed between two *lox* sites¹². To induce recombinational excision of the *RPM1* transgene, we crossed each replicate line with a Bla-2 line expressing *cre* recombinase and created selfed lines that had lost the *RPM1* transgene. This allowed us to generate independent pairs of homozygous lines with identical genetic backgrounds differing only with respect to the presence or absence of the *RPM1* gene. Individuals within each pair shared a common insertion site for the introduced DNA, but resistant individuals (*RPM1*⁺) expressed *RPM1* and the selectable marker, kanamycin, whereas susceptible individuals (*RPM1*[−]) expressed only kanamycin. The creation of paired lines effectively controls for fitness changes caused by the

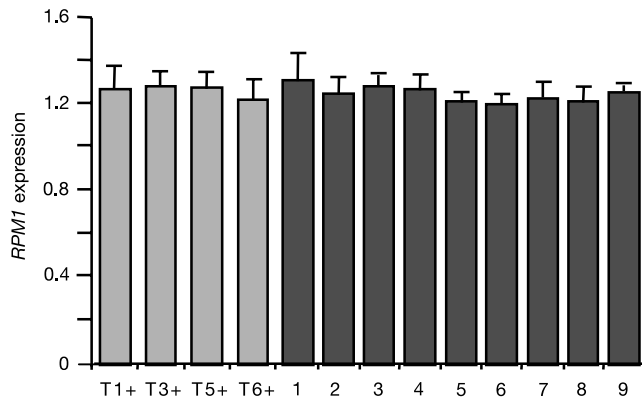


Figure 1 *RPM1* expression. Normalized expression levels ($RPM1$ threshold cycle (C_T) divided by endogenous C_T) in $RPM1^+$ transgenic plants (light grey bars) and nine accessions that possess *RPM1* naturally (dark grey bars; 1, Bur-0; 2, Col-0; 3, Ct-0; 4, Kas-1; 5, Lip-0; 6, Pog-0; 7, Tsu-0; 8, Tamm-7; 9, Wu-0). No *RPM1* expression was detected in the susceptible Bla-2 ecotype or the four $RPM1^-$ transgenic lines. Results are means $\pm$ s.e.m.

introduction of the foreign DNA (and a selectable marker) into each location in the genome.

We confirmed the appropriateness of our paired transgenic lines for use in fitness trials in six ways. First, we sequenced the transgenic *RPM1* alleles to confirm that they were unchanged. Second, we used anchor PCR¹³ to determine that each transgene landed in a non-coding DNA location with no known function. Third, we used direct sequencing to confirm that excisions occurred precisely at the *lox* sites, thus removing only the *RPM1* locus. Fourth, we verified that the $RPM1^+$, but not $RPM1^-$, lines elicited a hypersensitive response after treatment with *P. syringae* DC3000 carrying *AvrRpm1* (ref. 14). Fifth, we used quantitative real-time PCR to show that background (that is, uninduced) levels of *RPM1* expression in our transgenic lines were indistinguishable from expression levels of ecotypes carrying the *RPM1* allele (Fig. 1). Last, to further confirm that our transgenic lines did not overexpress *RPM1*, we tested for induction of the salicylic acid pathway. Our measurement of endogenous free salicylic acid and a defence induced by the salicylate-dependent pathway (peroxidase activity) failed to reveal any differences between the $RPM1^+$ and $RPM1^-$ lines (Table 1). Furthermore, growth of *P. syringae* pv. *tomato* DC3000 did not differ between $RPM1^+$ and $RPM1^-$ lines (Table 1). These results

indicate 'normal' functioning of the *RPM1* transgene in our experimental lines.

In late spring 2001, we transplanted 500 replicates of each line into the field, and grew them to maturity in the absence of detectable disease. To monitor for pathogenic bacteria, we collected and cultured bacteria from representatives of each plant line twice during plant growth. Although numerous bacterial isolates were collected, we found that they were not pathogenic and did not carry *AvrRpm1* or *AvrB*. When plants senesced, we counted the number of siliques per plant and the number of seeds per silique, and measured the dry biomass of each shoot. $RPM1^+$ plants tended to have fewer siliques and seeds per silique, as well as lower shoot biomass than the paired $RPM1^-$ plants (Table 2). More importantly, we found that $RPM1^+$ individuals suffered, on average, a 9% decrease in total seed production relative to their $RPM1^-$ counterparts (Fig. 2). There was no significant variation among the four pairs of transgenic lines in the magnitude of this cost of resistance (one-way analysis of variance on difference in total seed production between paired plants: $F_{3,1257} = 0.48$, $P > 0.6$). This consistency across independent transformed lines indicates that fitness loss was not a consequence of ectopic *RPM1* expression caused by position effects.

Given the large magnitude of this fitness cost, we asked whether the cost of resistance might have been an unintended consequence of inserting *RPM1* into a susceptible genetic background containing inappropriate alleles for an interacting cofactor. One such cofactor is *RIN4*, a gene that physically interacts with *RPM1* and is necessary for elicitation of a hypersensitive response¹⁵. Although *RIN4* and *RPM1* are not genetically linked, we hypothesized that, with strong epistasis, there might be permanent linkage disequilibrium between the two loci. We determined the *RPM1* genotype and the *RIN4* gene sequences of 96 ecotypes; no evidence for association between alleles at these two loci could be detected (Table 3). Indeed, Bla-2 contains a haplotype (type 3) of *RIN4* that is more commonly found in resistant individuals. Thus, we find no evidence for a mismatch between *RPM1* and *RIN4* alleles that could explain the fitness cost, but this does not rule out the possibility that other, as yet unidentified, cofactors are involved.

Two features of our experimental system facilitated the ability to detect a pleiotropic cost of *RPM1* resistance. First, we used genetic engineering to remove potentially confounding effects of linkage. Second, the susceptible allele, because it is a deletion of the entire locus, completely eliminates any possible direct or indirect cost of producing and expressing a 'susceptible' protein variant. Indeed, it has been suggested¹⁶ that previous failures to detect costs of R-gene resistance stemmed from a misclassification of 'defeated' (but

Table 1 Effect of *RPM1* presence or absence on the salicylate-dependent pathway

Line	Salicylic acid concentration (mmol per g dry leaf)			Peroxidase activity ($A_{470} \text{ min}^{-1}$ per mg total protein)			Bacterial growth (log colony-forming units)		
	$RPM1^+$	$RPM1^-$	<i>P</i>	$RPM1^+$	$RPM1^-$	<i>P</i>	$RPM1^+$	$RPM1^-$	<i>P</i>
T1	0.011 $\pm$ 0.001	0.011 $\pm$ 0.001	0.939	34.2 $\pm$ 20.0	34.3 $\pm$ 25.4	0.996	8.0 $\pm$ 0.3	7.6 $\pm$ 0.1	0.292
T3	0.012 $\pm$ 0.002	0.013 $\pm$ 0.001	0.674	7.7 $\pm$ 3.3	4.6 $\pm$ 1.0	0.887	7.5 $\pm$ 0.2	7.6 $\pm$ 0.1	0.832
T5	0.011 $\pm$ 0.001	0.009 $\pm$ 0.001	0.767	24.3 $\pm$ 9.5	28.8 $\pm$ 11.8	0.844	8.1 $\pm$ 0.4	7.8 $\pm$ 0.1	0.386
T6	0.012 $\pm$ 0.001	0.016 $\pm$ 0.006	0.412	44.4 $\pm$ 24.8	24.4 $\pm$ 11.6	0.388	7.9 $\pm$ 0.3	7.4 $\pm$ 0.6	0.211

Results are means $\pm$ s.e.m.

Table 2 Effect of *RPM1* presence or absence on correlates of fitness

Line	No. of siliques per plant			No. of seeds per silique			Shoot dry biomass (g)		
	$RPM1^+$	$RPM1^-$	<i>P</i>	$RPM1^+$	$RPM1^-$	<i>P</i>	$RPM1^+$	$RPM1^-$	<i>P</i>
T1	821.9 $\pm$ 32.6	889.6 $\pm$ 34.3	0.009	38.5 $\pm$ 0.4	40.5 $\pm$ 0.4	<0.001	1.23 $\pm$ 0.05	1.48 $\pm$ 0.05	<0.001
T3	670.3 $\pm$ 28.0	729.1 $\pm$ 28.3	0.008	38.2 $\pm$ 0.4	39.5 $\pm$ 0.4	0.151	1.21 $\pm$ 0.04	1.36 $\pm$ 0.05	0.001
T5	798.1 $\pm$ 29.6	835.6 $\pm$ 30.4	0.080	33.6 $\pm$ 0.3	35.4 $\pm$ 0.3	<0.001	0.94 $\pm$ 0.03	1.03 $\pm$ 0.04	0.003
T6	782.7 $\pm$ 31.9	802.9 $\pm$ 31.9	0.386	38.6 $\pm$ 0.4	41.1 $\pm$ 0.3	<0.001	1.31 $\pm$ 0.05	1.42 $\pm$ 0.05	0.015

Results are means $\pm$ s.e.m.

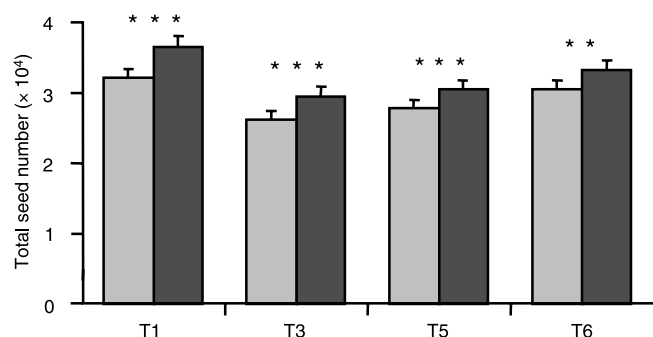


Figure 2 Total seed numbers. Seed production of *RPM1*⁺ sibs (light grey bars) and *RPM1*⁻ sibs (dark grey bars) in each transgenic line (T1, T3, T5, T6) is shown. Total seed number was estimated as the number of siliques multiplied by the average seed number per silique ($n = 5$). Results are means $\pm$ s.e.m. Data were analysed with paired *t*-tests for each background. Two asterisks, $P < 0.01$; three asterisks, $P < 0.001$.

functional) resistance alleles as susceptibility alleles. The issue of whether R-genes have a cost of resistance has been actively debated in the literature^{1,3,17–19}. This is the first definitive demonstration that such a cost exists. In addition, we find that the magnitude of this cost is surprisingly large, easily exceeding the 3.5% average fitness cost of disease resistance revealed by a meta-analysis of published studies of costs of resistance²⁰.

The cost of *RPM1* resistance is unlikely to be explained by the metabolic cost of *RPM1* synthesis, as the constitutive level of *RPM1* is extremely low¹⁰. Although we currently lack a molecular mechanism to explain the reduction in the fitness of resistant plants, two hypotheses can be proposed. First, *RPM1*⁺ plants might trigger the induction of plant defence pathways because of *RPM1* overexpression in the absence of pathogens. Although *RPM1* was expressed at normal levels in our greenhouse plants (and the salicylate-dependent defence pathway does not seem to have been induced), recent work on the rice-blast-resistance gene family demonstrates that R-gene expression levels can be markedly upregulated by environmental signals²¹. If this occurred during our field trial, then fitness could decline because R-gene overexpression can lead to constitutive defence response^{22,23}, cell death^{23–25} and even plant death²⁶. The second hypothesis is that basal levels of *RPM1* indirectly induce plant defence responses as a consequence of interactions with other R-proteins, the targets they guard, and the bacterial effectors (Avr proteins) they recognize^{15,27–29}. Distinguishing between these alternative hypotheses will be important for understanding R-gene structure/function and evolution.

With more than 100 R-gene loci spread across the *Arabidopsis* genome, many of which are likely to be segregating for resistance and susceptibility alleles, can the fitness cost of *RPM1* resistance be representative of other R-genes? We think not. First, the presence of so many loci segregating for alleles with 9% fitness costs would impose an impossibly large genetic load on populations. Second,

the fact that *RPM1* segregates for a long-lived polymorphism for resistance and susceptibility alleles could bias it towards having a large fitness cost of resistance. Theoretical models predict that resistance alleles with the largest costs are least likely to be driven to fixation in the species, and thus facilitate long-term maintenance of polymorphism³⁰. Our previous molecular evolutionary analyses of the *RPM1* flanking sequence suggested a trench warfare model in which resistance and susceptibility alleles fluctuate in response to pathogen epidemiology but are both maintained because of the costs and benefits of resistance³. The recent demonstration that pseudomonads attack *A. thaliana* in natural populations³¹ and the current demonstration of a substantial cost of resistance lend further support to this model. □

Methods

Cre-lox insertion of *RPM1*

To create *lox-RPM1-lox* lines, we cloned a 3.84-kb region from Columbia, containing *RPM1*, its minimal promoter and its full terminator, between two *lox* sites in the vector pBS246 (Gibco-BRL catalogue no. 10349-019; Life Technologies). There is relatively little non-coding DNA between *RPM1* and adjacent loci (577 base pairs (bp) 5' and 989 bp 3'), and our transgene contains almost all of this sequence. Thus, unless *RPM1* cis-regulation extends into or across the adjacent loci, our transgene contains the entire cis-regulatory region. A 4-kb fragment with *lox* sequences on either side of *RPM1* was then cloned into the binary vector pBin19. This construct, which included the selectable marker *nptII* outside the *lox* sites, was introduced by means of vacuum infiltration into Bla-2.

To create *cre* lines, a 3.43-kb region containing *cre* with the 35S promoter and Nos3' terminator, as well as the selectable markers Basta and GUS, was cloned in the binary vector pCambia3301. This construct was used to create six *cre* lines of the susceptible ecotype Bla-2. The *lox-RPM1-lox* lines were emasculated and then pollinated by the *cre* lines. In the F₂ generation, the presence of *cre* was negatively selected with Basta (AgroEvo USA) diluted 1:1,100. The remaining plants were screened by PCR to identify an individual that had lost *RPM1* and its sib that had retained *RPM1*. After confirmation that these lines were not resistant to GUS or Basta, each pair was self-pollinated for five generations and the homozygosity of *RPM1*⁺ or *RPM1*⁻ was confirmed by PCR. Inserts were sequenced in each *RPM1*⁺ line to confirm their integrity, and in each *RPM1*⁻ line to confirm clean excision. Southern blots and anchor PCR tests were consistent with the presence of only a single insert for each transformed line.

Quantitative real-time PCR

Eight PCR replicates from three independent RNA extractions were run for each line on an ABI Prism 7700 sequence-detection system with TaqMan PCR core reagent (PE Biosystems). Leaves were sampled from 4-week-old plants. EF1a, an elongation factor, was used as an endogenous reference gene in the same reaction tube³².

Field experiment

Seedlings of each *RPM1*⁺ and *RPM1*⁻ line were germinated in plug trays containing Premier Pro-Mix in the University of Chicago greenhouse. At the four-leaf stage, 4,000 seedlings (500 per line) were transplanted into a field site in Downer's Grove, Illinois, in a pseudo-random design in which *RPM1*⁺ and *RPM1*⁻ sibs were paired, and lines were cycled throughout the field. Plants were set out in 20 rows of 200 each, spaced by 0.3 m within rows and by 1 m between rows. Plants were irrigated for 1 week to reduce transplantation shock, and then were sustained only by natural rainfall. Plants were hand-weeded once and received no protection from pests. Roughly 21% of plants died, but these were evenly distributed among the treatments. Pairs of plants were excluded from analysis if either individual died prematurely.

Field assay of bacteria

Ninety-six leaf samples (representing plants of all eight lines) were ground and plated on KB medium on two dates during the field experiment. Bacteria were found in 24 samples from day 14, and in 32 samples from day 30. Out of hundreds of colonies, 15 potentially distinct bacterial types were identified by general appearance (colour, size and growth pattern). Each type was tested for pathogenicity on tobacco at a concentration of 10⁸ colony-forming units per ml (ref. 33), sequenced for 16S rRNA (ref. 34) and compared to published sequences in GenBank. None of these bacterial isolates were pathogens, according to the tobacco HR test or their 16S identity. Finally, we used PCR to screen for the presence of *AvrRpm1* and *AvrB*. For *AvrRpm1*, we designed a pair of degenerate primers for conserved regions of the gene based on three homologous sequences in GenBank. We also screened with a pair of primers that reliably amplified the *P. syringae* gene. For *AvrB*, only a single published sequence was available. We screened with nine combinations of primer pairs based on the published sequence. We failed to detect the *AvrRpm1* or *AvrB* genes in any of our colonies.

Chemical analyses

Five replicates of each line were grown in the greenhouse and harvested at 3 weeks for chemical analysis. Leaf material was frozen in liquid nitrogen, freeze-dried and homogenized. Material for salicylic acid analysis was extracted three times with 70% methanol and quantified by high-performance liquid chromatography³⁵. Material for protein analysis was extracted in 0.25 ml of 0.05 M sodium phosphate buffer pH 7.0.

Table 3 Common haplotypes (more than two individuals) in *RIN4* gene among 96 *A. thaliana* accessions

Type	No. of accessions*	Amino acid location						Resistance genotype	
		38	56	89	129	134	147	<i>RPM1</i> ⁺	<i>RPM1</i> ⁻
1	41	–	–	–	–	–	–	30	11
2	17	V	–	–	T	A	–	14	3
3	14	–	P	–	T	A	–	12	2
4	11	–	–	D	–	–	–	10	1
5	5	–	–	–	T	A	–	4	1
Consensus		M	H	G	S	T	V	70	18

*Eight accessions exhibited rare haplotypes.

Peroxidase activity was determined by following the oxidation of guaiacol for 1 min at 470 nm and was standardized by total protein content as previously described³⁶.

Bioassay with *P. syringae* pv. *tomato* DC3000

Fifteen replicates of each line were challenged by infiltration of three leaves with 10⁴ colony-forming units per ml ($D_{600} = 0.002$) after 3 weeks of plant growth. After 5 days, leaf discs were removed, ground and plated on KB medium to determine the concentration of bacteria.

Received 18 February; accepted 18 March 2003; doi:10.1038/nature01588.

- Grant, M. R. *et al.* Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **87**, 15843–15848 (1998).
- Caicedo, A. L., Schaal, B. A. & Kunkel, B. N. Diversity and molecular evolution of the *RPS2* resistance gene in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **96**, 302–306 (1999).
- Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667–671 (1999).
- Tian, D., Araki, H., Stahl, E., Bergelson, J. & Kreitman, M. Signature of balancing selection in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11525–11530 (2002).
- Mauricio, R., Korves, T., Stahl, E. A., Kreitman, M. & Bergelson, J. Natural selection for polymorphism in the disease resistance gene *RPS2* of *Arabidopsis*. *Genetics* **163**, 735–746 (2003).
- Simms, E. L. & Rausher, M. D. Costs and benefits of plant resistance to herbivory. *Am. Nat.* **130**, 570–581 (1987).
- Bergelson, J., Dwyer, G. & Emerson, J. J. Models and data on plant–enemy coevolution. *Annu. Rev. Genet.* **35**, 469–499 (2001).
- Boyes, D. C., Nam, J. & Dangel, J. L. The *Arabidopsis thaliana* *RPM1* disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proc. Natl Acad. Sci. USA* **95**, 15849–15854 (1998).
- Bisgrove, S. R., Simonich, M. T., Smith, N. M., Sattler, A. & Innes, R. W. A disease resistance gene in *Arabidopsis* with specificity for two different pathogen avirulence genes. *Plant Cell* **6**, 927–933 (1994).
- Grant, M. R. *et al.* Structure of the *Arabidopsis* *RPM1* gene enabling dual specificity disease resistance. *Science* **269**, 843–846 (1995).
- Redei, G. P. *Arabidopsis* as a genetic tool. *Annu. Rev. Genet.* **9**, 111–127 (1986).
- Ow, D. W. & Medberry, S. L. Genome manipulation through site-specific recombination. *Crit. Rev. Plant Sci.* **14**, 239–261 (1995).
- Howe, C. *Gene Cloning and Manipulation* (Cambridge Univ. Press, Cambridge, 1995).
- Whalen, M. C., Innes, R. W., Bent, A. F. & Staskawicz, B. J. Identification of *Pseudomonas syringae* pathogens of *Arabidopsis* and a bacterial locus determining avirulence on both *Arabidopsis* and soybean. *Plant Cell* **3**, 49–59 (1991).
- Mackey, D., Holt, B. E., Wiig, A. & Dangel, J. L. RIN4 interacts with *Pseudomonas syringae* Type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* **108**, 743–754 (2002).
- Jorgensen, J. H. & Jensen, H. P. Effect of ‘unnecessary’ powdery mildew resistance genes on agronomic properties of spring barley. *Norsk Landbruksforskning. Suppl.* **9**, 125–130 (1990).
- Brown, J. K. M. Yield penalties of disease resistance in crops. *Curr. Opin. Plant Biol.* **5**, 1–6 (2002).
- Rausher, M. D. Co-evolution and plant resistance to natural enemies. *Nature* **411**, 857–864 (2001).
- Stuiver, M. H. & Custers, J. H. H. V. Engineering disease resistance in plants. *Nature* **411**, 865–868 (2001).
- Bergelson, J. & Purrington, C. B. Surveying patterns in the cost of resistance in plants. *Am. Nat.* **148**, 536–558 (1996).
- Wang, Z.-W., Yamanouchi, U., Katayose, Y., Sasaki, T. & Yano, M. Expression of the *Pib* rice-blast-resistance gene family is up-regulated by environmental conditions favouring infection and by chemical signals that trigger secondary plant defences. *Plant Mol. Biol.* **47**, 653–661 (2001).
- Oldroyd, G. E. D. & Staskawicz, B. J. Genetically engineered broad-spectrum disease resistance in tomato. *Proc. Natl Acad. Sci. USA* **95**, 10300–10305 (1998).
- Tang, X. *et al.* Overexpression of *Pto* activates defense responses and confers broad resistance. *Plant Cell* **11**, 15–29 (1999).
- Mindrinos, M., Katagiri, F., Yu, G. & Ausubel, F. M. The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* **78**, 1089–1099 (1994).
- Tao, Y., Yuan, F., Leister, R. T., Ausubel, F. M. & Katagiri, F. Mutational analysis of the *Arabidopsis* nucleotide binding site-leucine-rich repeat resistance gene *RPS2*. *Plant Cell* **12**, 2541–2554 (2000).
- Holt, B. J. *et al.* An evolutionarily conserved mediator of plant disease resistance gene function is required for normal *Arabidopsis* development. *Dev. Cell* **2**, 807–817 (2002).
- van der Biezen, E. A. & Jones, J. D. G. Plant disease resistance proteins and the gene for gene concept. *Trends Biochem. Sci.* **23**, 454–456 (1998).
- Dangel, J. L. & Jones, J. D. G. Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–832 (2001).
- Mackey, D., Belkadir, Y., Alonso, J. M., Ecker, J. R. & Dangel, J. L. *Arabidopsis* RIN4 is a target of the Type III virulence effector *AvrRpt2* and modulates *RPS2*-mediated resistance. *Cell* **112**, 379–389 (2003).
- Gillespie, J. H. Natural selection for resistance to epidemics. *Ecology* **56**, 493–495 (1975).
- Jakob, K. *et al.* *Pseudomonas viridiflava* and *P. syringae*—natural pathogens of *Arabidopsis thaliana*. *Mol. Plant–Microbe Interact.* **15**, 1195–1203 (2002).
- Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6**, 986–994 (1996).
- Klement, Z. Rapid detection of the pathogenicity of phytopathogenic *Pseudomonads*. *Nature* **199**, 299–300 (1963).
- Moore, E. R. B. *et al.* The determination and comparison of the 16S rRNA gene sequences of species of the genus *Pseudomonas* (sensu stricto) and estimation of the natural intrageneric relationships. *Syst. Appl. Microbiol.* **19**, 478–492 (1996).
- Dewdney, J. *et al.* Three unique mutants of *Arabidopsis* identify *eds* loci required for limiting growth of a biotrophic fungal pathogen. *Plant J.* **24**, 205–218 (2000).
- Traw, M. B., Kim, J., Enright, S., Cipollini, D. F. & Bergelson, J. Negative cross-talk between the salicylate and jasmonate-mediated pathways in the Wassilewskija ecotype of *Arabidopsis thaliana*. *Mol. Ecol.* **12**, 1125–1135 (2003).

Acknowledgements We thank J. Dangel for feedback, and members of the Department of Ecology and Evolution for assistance in the field. This research was supported by NIH grants to J.B.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.B. (jbergels@midway.uchicago.edu).

Non-classical receptive field mediates switch in a sensory neuron’s frequency tuning

Maurice J. Chacron^{*†}, Brent Doiron^{*†}, Leonard Maler[†], André Longtin^{*‡} & Joseph Bastian[‡]

^{*} Physics Department, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5

[†] Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario, Canada K1H 8M5

[‡] Department of Zoology, University of Oklahoma, Norman, Oklahoma 73019, USA

Animals have developed stereotyped communication calls to which specific sensory neurons are well tuned^{1,2}. These communication calls must be discriminated from environmental signals such as those produced by prey. Sensory systems might have evolved neural circuitry to encode both categories. In weakly electric fish, prey and communication signals differ in their spatial extent and frequency content^{3,4}. Here we show that stimuli of different spatial extents mimicking prey and communication signals cause a switch in the frequency tuning and spike-timing precision of electrosensory pyramidal neurons, resulting in the selective and optimal encoding of both stimulus categories. As in other sensory systems⁵, pyramidal neurons respond only to stimuli located within a restricted region of space known as the classical receptive field (CRF)⁶. In some systems, stimulation outside the CRF but within a non-classical receptive field (nCRF) can modulate the neural response to CRF stimulation even though nCRF stimulation alone fails to elicit responses^{7,8}. We show that pyramidal neurons possess a nCRF and that it can modulate the response to CRF stimuli to induce this neurobiological switch in frequency tuning.

The complex statistical structure of many naturalistic visual⁹ and auditory¹⁰ stimuli makes our interpretation of neural responses to these stimuli difficult and often prevents clearcut correlations of the responses with behaviour. Weakly electric fish offer a simple system for studying the differential encoding of natural stimuli because there is a clear spatiotemporal distinction between prey and communication stimuli. Amplitude modulations (AMs) of the electric fish’s self-generated electric organ discharge (EOD) contain information relevant to both types of stimulus¹¹. Epidermal electroreceptors encode these AMs precisely¹² and provide synaptic input¹³ to pyramidal neurons of the electrosensory lateral line lobe (ELL), whose antagonistic centre-surround CRF structure⁶ resembles that of visual neurons⁵. Relative motion of the fish near prey during feeding produces low-frequency (less than 10 Hz) spatially localized AMs³. However, communication signals from conspecifics produce high-frequency (more than 50 Hz) spatially diffuse AMs⁴.

To provide naturalistic stimuli that mimic prey and communication signals, we used two stimulation geometries (see Methods). Local stimulus geometry provides AMs whose spatial extent is similar to that produced by prey, while global stimulus geometry produces spatially diffuse AMs similar to communication signals

(Fig. 1a, b). We investigated the frequency response properties of ELL pyramidal neurons by constructing EOD AM frequency tuning curves. Changing the stimulus geometry from local to global causes a shift in response preference from low to high frequencies (Fig. 1c; compare blue with red). We quantified this shift by computing the

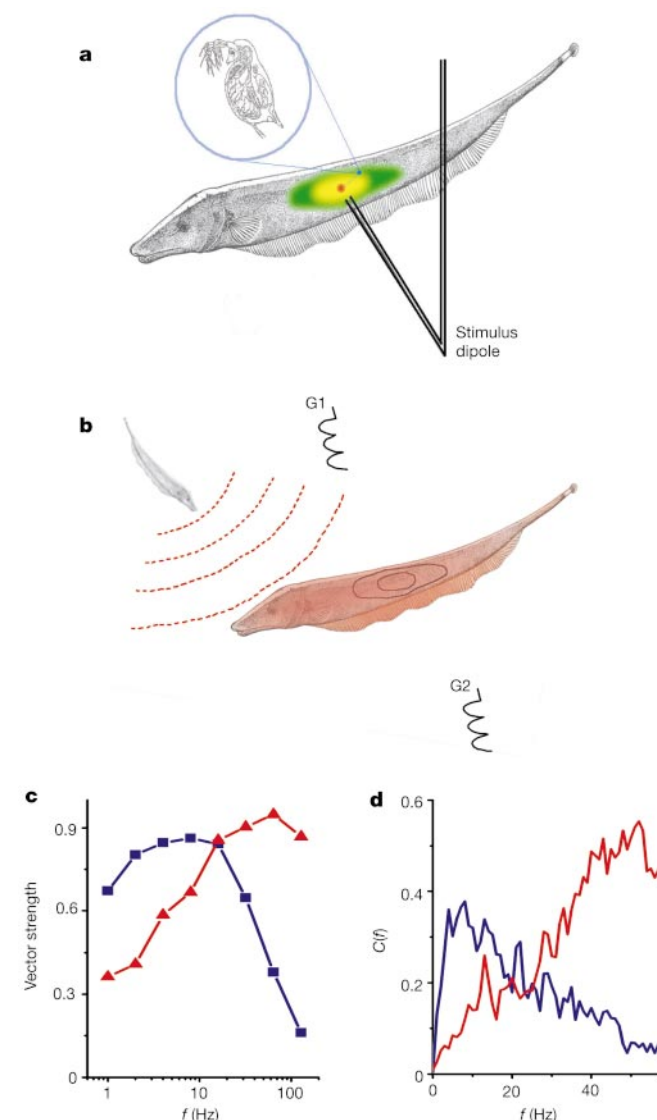


Figure 1 ELL pyramidal neurons display differential frequency tuning to local and global stimulation geometries. **a**, Local stimulation geometry. A small stimulus dipole (2-mm tip spacing) was positioned 2–3 mm lateral to the fish and delivered AMs of the animal's own electric organ discharge. These AMs were spatially localized (red circle) over only a fraction of the CRF centre (yellow ellipse). They did not stimulate the surround (green ellipse) or the nCRF lying outside the surround. The electric image was spatially similar to that produced by a prey (*Daphnia*, blue circle) as the fish swims by. **b**, Global stimulation geometry. Two silver–silver-chloride electrodes (G1 and G2) were placed about 19 cm lateral to each side of the animal. Global AMs (red) stimulate the entire body surface including the CRF centre and surround and also any putative nCRF, and are spatially similar to those produced by communication stimuli. In addition, the ipsilateral and contralateral body surfaces receive stimuli in antiphase. **c**, Tuning curves (vector strength plotted against frequency) for single-unit responses from ELL pyramidal neurons obtained under local (blue) and global (red) stimulus geometries. The vector strength under local stimulus geometry is maximal for frequencies of less than 10 Hz. However, it is maximal for frequencies of about 100 Hz under global stimulus geometry. **d**, Coherence curve (see Methods) obtained under local (blue) and global (red) stimulus geometries. The curves are qualitatively similar to the tuning curves obtained in **c**.

frequencies associated with the maximum vector strength (see Methods) under each stimulus geometry. These averaged 4.67 ± 2.73 Hz and 53.33 ± 16.52 Hz, respectively (mean change of 48.66 Hz, $P = 0.0004$, pairwise t -test, $n = 7$). Plots of the stimulus–spike train coherence obtained from random AM stimulation (see Methods) showed a qualitatively similar effect (Fig. 1d; compare blue with red). The average coherence values C_{low} between 0 and 20 Hz and C_{high} between 40 and 60 Hz were also computed. With global stimulus geometry, high-frequency coherence was significantly greater than low-frequency coherence ($P < 10^{-3}$, t -test, $n = 17$), but with local stimulus geometry the opposite frequency preference was seen ($P < 10^{-3}$, t -test, $n = 17$). Firing rates did not change as we went from local (23.12 ± 10.86 spikes s^{-1}) to global (23.24 ± 11.82 spikes s^{-1}) geometry ($P = 0.9$, pairwise t -test, $n = 17$). This confirms that ELL pyramidal neurons behave as low-pass filters when stimulated locally, whereas they exhibit high-pass characteristics when stimulated globally.

We used information theoretic measures^{14,15} (see Methods) to quantify the consequences for stimulus encoding of this shift in temporal frequency tuning. We compared results obtained with locally and globally applied low-frequency (0–20-Hz) and high-frequency (40–60-Hz) random AMs. These results are summarized in Table 1. High-frequency global stimuli (communication-like) were encoded much better (about 200%) than high-frequency local stimuli. Moreover, low-frequency local stimuli (prey-like) were encoded much better (about 100%) than low-frequency global stimuli. These results demonstrate that pyramidal neurons show improved stimulus encoding when the combination of temporal frequency content and spatial stimulus characteristics closely mimic those of natural stimuli.

The higher mutual information rates, indicating an increased signal-to-noise ratio, obtained with high-frequency global stimuli as compared with high-frequency local stimuli indicate a possible dependence of spike train variability on the spatial extent of the stimulus. Neurons can display low trial-to-trial variability to repeated stimuli both *in vitro*¹⁶ and *in vivo*¹⁷. Using ‘frozen noise’ (see Methods), we explored the reliability of spike timing displayed by ELL pyramidal neurons. Results show high trial-to-trial variability under local stimulus geometry (Fig. 2a) and low trial-to-trial variability under global stimulus geometry (Fig. 2b). The average reliability of spike timing, measured as the fraction of spikes occurring reproducibly during high-frequency ‘events’¹⁶, increased from 0.16 ± 0.17 under local stimulation to 0.72 ± 0.12 with global stimulation ($P = 0.004$, pairwise t -test, $n = 7$). The mean spike time precision (average standard deviation of spike times within ‘events’¹⁶) decreased from 1.44 ± 0.16 ms under local stimulation to 1.08 ± 0.23 ms with global stimulation ($P = 0.009$, pairwise t -test, $n = 7$). Pyramidal neurons display high reliability and high spike timing precision to stimuli mimicking the frequency content of communication signals, but only when the stimulus is presented globally. Such timing precision does not occur for low-frequency stimuli (data not shown), presumably because accurate encoding of low-frequency stimuli does not require such high temporal precision.

Table 1 Summary of results with different information-theoretic measures

	Local	Global
High frequency	$\gamma = 0.08 \pm 0.08$ $I = 0.29 \pm 0.28$	$\gamma = 0.25 \pm 0.17$ $I = 1.14 \pm 0.52$
Low frequency	$\gamma = 0.26 \pm 0.11$ $I = 0.74 \pm 0.33$	$\gamma = 0.15 \pm 0.12$ $I = 0.36 \pm 0.20$
Statistical significance	$P_{\gamma} = 0.0031$ ($n = 12$) $P_I = 0.0035$ ($n = 12$)	$P_{\gamma} = 0.0015$ ($n = 11$) $P_I = 0.0014$ ($n = 11$)

The coding fraction, γ , is the fraction of the stimulus waveform correctly estimated by the neuron; the mutual information rate, I , in bits per spike gives the amount of information transmitted by the neuron. P_{γ} and P_I are the respective P values obtained from a pairwise t -test comparing the γ and I values obtained with high-frequency (40–60-Hz) stimuli with those obtained with low-frequency (0–20-Hz) stimuli.

Anatomical studies^{18,19} predict that local and global stimulation will activate different constellations of synaptic inputs to ELL pyramidal neurons. If these differing synaptic inputs cause the shift in frequency tuning, this should be reflected in the cell's membrane potential. We recorded intracellularly from pyramidal neurons to test this hypothesis. Figure 2 shows the membrane potential response under local (Fig. 2c) and global (Fig. 2d) stimulation geometry with random AM stimulation. The membrane potential response (black) does not track the higher-frequency components of the stimulus (blue) during local stimulation but does so during global stimulation. We emphasize this point by computing the coherence between the membrane potential response with spikes removed (green) (see Methods) and the stimulus (Fig. 2e, f). For local stimulus geometry, C_{low} was significantly greater than C_{high} ($P = 0.024$, t -test, $n = 6$) whereas the opposite was true for global stimulus geometry ($P = 0.012$, t -test, $n = 6$). These results show that the change in frequency

tuning is seen in the membrane potential response itself and thus probably originates from differing synaptic inputs under local and global stimulus geometries.

We have previously mapped the antagonistic centre-surround CRF organization of ELL pyramidal neurons⁶. Local stimuli affect only a fraction of the CRF centre (Fig. 1a) whereas global stimuli influence the entire CRF as well as the nCRF (Fig. 1b). In the visual system, the nCRF is known to modulate CRF centre responses^{7,8}. To test for the presence of nCRF effects and their possible role in pyramidal neuron frequency tuning shifts, we performed partition experiments (see Methods) illustrated in Fig. 3a. The thin rubber partition electrically isolated the fish's head region from its trunk region, allowing each to be independently stimulated. We recorded from pyramidal neurons whose CRF centre was sufficiently distant from the partition to ensure that the CRF was entirely within the trunk region. Thus, stimuli applied to the head region influenced the responsiveness of the recorded cell only through the nCRF. Local stimulation of the CRF centre alone with 0–60 Hz random AMs produced results identical to those obtained under local stimulus geometry without a partition (compare Figs 1d and 3b, blue). We then paired the local CRF centre stimulation with in-phase global stimulation of the head chamber (Fig. 3b, red). Simultaneous stimulation of the nCRF decreased the cell's response to low frequencies only. The measure of low-frequency coherence, C_{low} , was significantly decreased ($P < 10^{-3}$, pairwise t -test, $n = 15$). This decrease is similar to that seen in the transition from local to global geometries. However, nCRF stimulation phase-shifted by 180° relative to the CRF had no effect on coherence (Fig. 3b, green); there was

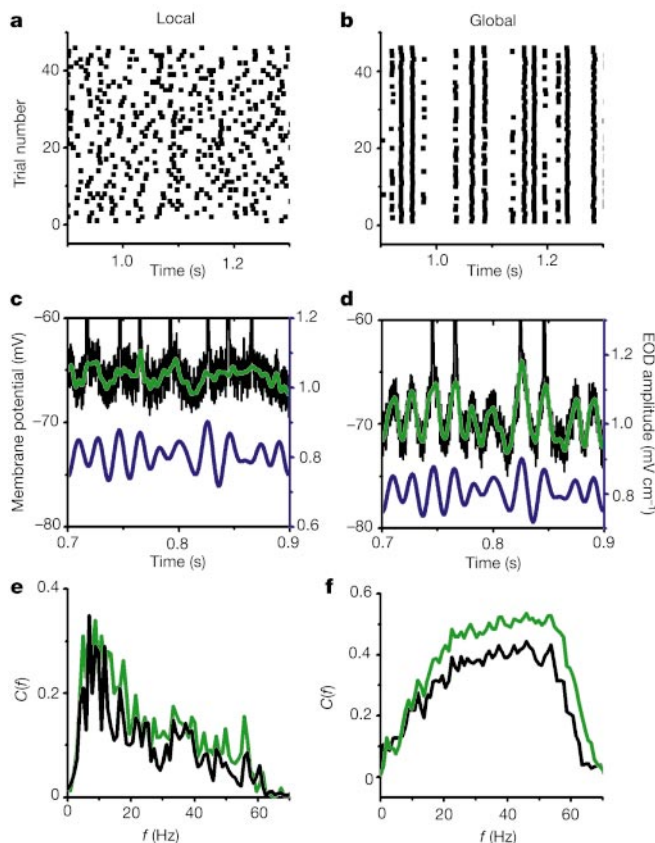


Figure 2 Spike timing reliability and precision under local and global stimulation geometries. **a**, Raster plot showing responses to short subsets of 2-s epochs of frozen 40–60-Hz random AMs presented locally. **b**, As in **a**, but the stimuli were presented with global stimulus geometry. **c**, Intracellularly recorded membrane potential response V_m (black trace) from a pyramidal cell under local 0–60-Hz random AM stimulation. The ~ 72 -mV-high spikes are truncated. The green trace shows V_m after spikes had been removed and replaced with the means of V_m immediately before and after each spike, then low-pass filtered (200-Hz cut-off, eighth-order Butterworth). The blue trace shows the time course of EOD AMs measured within the centre of the cell's receptive field and low-pass filtered with the same filter parameters. The EOD AM was shifted by ~ 7 ms to compensate for axonal and synaptic delays. Comparison of the green and blue traces shows that V_m fails to follow the high-frequency components of the stimulus. **d**, V_m from the same cell but with global stimulus geometry. **e**, Coherence between stimulus and membrane potential response with spikes removed (green) and coherence between stimulus and spike train (black) under local 0–60-Hz random AM stimulation. **f**, As in **e**, but with global stimulus geometry.

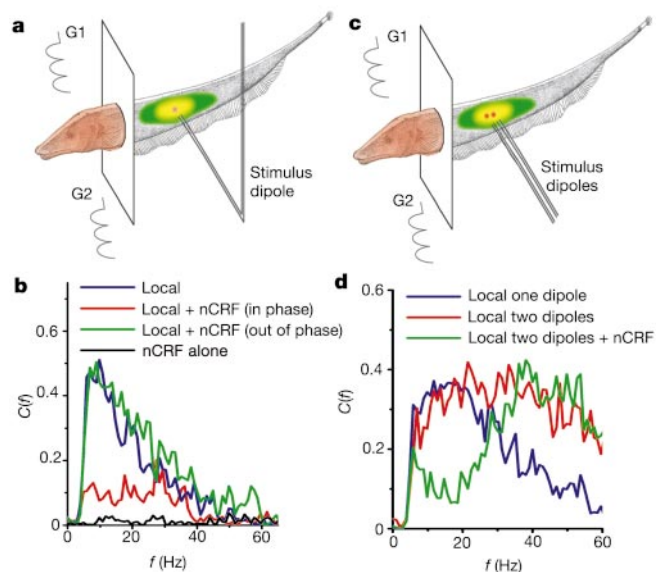


Figure 3 The nCRF mediates pyramidal neuron frequency tuning. **a**, Partition experiments. The animal's head was electrically isolated from the trunk by a thin rubber partition so that stimuli could be presented to the head and trunk regions independently (see Methods). **b**, Effects of pairing local trunk stimulation with global stimulation of the head chamber. Local stimulation of the CRF centre alone produces high coherence at low frequencies and low coherence at high frequencies (blue), as in Fig. 1c. Pairing local stimulation with in-phase head-chamber (nCRF) stimulation decreased low-frequency responsiveness (red); 180° phase-shifting of the nCRF stimulus relative to the CRF centre stimulus resulted in no loss of low-frequency coherence (green). nCRF stimulation alone was ineffective (black). **c**, Increased spatial saturation of the CRF centre was accomplished with two stimulus dipoles. **d**, Stimulation of the CRF centre by two dipoles (red trace) preferentially increased the cell's response to high frequencies over stimulation with one dipole (blue trace). The addition of nCRF stimulation attenuated low-frequency responsiveness (green).

no significant change in C_{low} and C_{high} ($P = 0.36$ and 0.16 respectively, pairwise t -tests, $n = 15$). Stimulation of the head region alone was ineffective in driving the recorded cell (Fig. 3b, black).

The cells' firing rates decreased by 3.9 and 2.1 spikes s^{-1} on average under in-phase and out-of-phase paired stimulation of CRF plus nCRF, respectively ($P = 0.003/0.001$, pairwise t -tests, $n = 15$). Because, in each case, the neurons received the same amount of primary afferent excitation by means of the CRF centre, the cells must be inhibited by nCRF stimulation. Furthermore, because the nCRF stimulus must be in phase with the CRF stimulus to decrease a neuron's response to low frequencies, the inhibition must act on a moderately fast timescale. Paired stimulation of the CRF centre plus the nCRF in phase did not on average alter high-frequency response because C_{high} did not change significantly with in-phase nCRF stimulation ($P = 0.58$, pairwise t -test, $n = 15$).

Either global stimulus geometry or paired stimulation of the nCRF plus the CRF centre in the partition experiment reduced low-frequency coherence. However, only global stimulation (no partition) resulted in an improvement in high-frequency coherence. Global geometry not only results in nCRF stimulation but also ensures that the entire CRF centre is stimulated (spatial saturation). Hence, the complete shift in frequency tuning seen with the transition from local to global geometry (Fig. 1c, d) might require both stimulation of the nCRF as well as spatial saturation of the CRF centre. To test the latter hypothesis we repeated the partition experiment but increased the CRF centre area influenced by the local stimulus by adding a second dipole (Fig. 3c; see Methods). As shown in Fig. 3d (compare blue and red), addition of the second dipole increased C_{high} by about 70% ($P = 0.0027$, pairwise t -test, $n = 7$). Thus, combining saturation of the CRF centre with in-phase nCRF stimulation is sufficient to induce both the increase in the cell's high-frequency response and the decrease in its low-frequency response (Fig. 3d, green) just as in the normal transition from local to global geometry. Stimulation of the CRF surround did not alter the frequency tuning to stimulation of the CRF centre (data not shown).

We have demonstrated that ELL pyramidal neurons can switch their tuning properties on the basis of the spatial extent of a stimulus. Specifically, pyramidal neurons responded maximally to temporal frequencies below 10 Hz under spatially local stimulation and to frequencies over 50 Hz under spatially global stimulation. These responses are well matched to the observed temporal frequency content of prey³ and communication⁴ stimuli, respectively. Maximum information transfer was obtained when the stimulus matched the spatiotemporal content of prey and communication stimuli. In particular, high information transfer was obtained for high-frequency global stimuli. This differs from previous results^{6,12} that showed poor information transfer under a global-like geometry. However, both previous studies used mainly low-frequency stimuli, to which pyramidal neurons respond poorly under global stimulation.

Intracellular recordings revealed that a change in synaptic input was probably responsible for the change in frequency tuning. The reduced response to low frequencies under global stimulation is due primarily to nCRF stimulation, probably acting through inhibition because a decrease in firing rate was observed. This decreased response is thus unlikely to occur with global yet spatially heterogeneous environmental stimuli such as those caused by a root mass²⁰, because it requires spatially homogeneous communication-like signals. ELL pyramidal neurons receive inhibition from several sources^{18,19}. Inhibition is known to modulate neural frequency tuning²¹ and can lead to oscillations and synchrony in a neural population²². We have recently shown that ELL pyramidal neurons displayed inhibition-mediated oscillatory dynamics under global stimulation but not under local stimulation²³. The emergence of this oscillation can sometimes accompany the shift in frequency response found here, but is not required to induce it (data not shown).

Our study shows that the structure of the receptive field of ELL pyramidal neurons is well adapted to categorical coding of their natural stimulus environment. This is achieved through the differential spatial extent of prey and communication stimuli that will differentially activate the structure of the receptive field of pyramidal neurons. Prey-like stimuli activate only a fraction of the CRF centre, whereas communication stimuli spatially saturate the CRF centre and also activate the nCRF, thus producing different synaptic input from that for prey stimuli. This neurobiological switch in synaptic input allows pyramidal neurons to encode both prey and communication stimuli optimally. □

Methods

Stimulation and recording

The experimental protocol has been described previously⁶. The stimuli consisted of sinusoidal and band-limited random AMs of an animal's own EOD presented with local or global geometry. When two local stimulus dipoles were used, the same stimulus waveform was fed to independent stimulus isolation units, each of which drove one dipole. Recordings were limited to E-type pyramidal neurons of the contralateral and lateral segments^{24,25}, which are important for processing both high-frequency communication signals^{26,27} and prey stimuli⁶. The centromedial ELL segment, required for the jamming avoidance response evoked by low-frequency global stimuli²⁸, is not considered here.

Extracellular single-unit recordings were made with metal-filled microelectrodes, and intracellular recordings were made with 40–100-M Ω KCl-filled micropipettes. Standard methods of preamplification were used and data were acquired with Cambridge Electronic Design 1401plus hardware and SpikeII software. All surgical procedures were performed in accordance with the University of Oklahoma animal care and use guidelines.

Data analysis

Spike trains during sinusoidal AM stimulation were accumulated as cycle histograms and the response was quantified by using the vector strength^{6,29}, which measures the degree of phase locking and ranges between 0 (no phase locking) and 1 (perfect phase locking).

Responses to random AMs were analysed by computing the coherence, $C(f)$, between the spike train and stimulus, where $C(f) = |P_{sx}(f)|^2 / [P_{ss}(f)P_{xx}(f)]$. $P_{ss}(f)$ and $P_{xx}(f)$ are the power spectra of the stimulus and spike train respectively, and $P_{sx}(f)$ is the cross-spectrum between the stimulus and the spike train. $C(f)$ ranges from 0 to 1 and indicates the strength of the response to the stimulus at a frequency f . Animals often displayed electrocommunication responses to random AMs, but only when these were applied globally (data not shown), indicating that these stimuli are good communication signal mimics.

We used the 'frozen noise' technique to examine trial-to-trial variability: the same random AM (40–60 Hz, duration 2 s) was delivered at least 30 times. We computed the reliability and precision measures from post-stimulus time histograms (binwidth 3 ms) as described previously¹⁶.

Information theoretic measures

A lower bound on the mutual information rate (in bits s^{-1}) is given in refs 14 and 30:

$$I = - \int_{f_{\text{low}}}^{f_{\text{high}}} df \log_2 [1 - C(f)]$$

where f_{low} and f_{high} define the bandwidth of the stimulus. We obtained the mutual information rate in bits per spike by dividing I by the cell's mean firing rate during stimulation.

We performed linear stimulus reconstruction^{14,30} as reported previously⁶. Note that this provides an absolute lower bound on information transmission³⁰. The quality of the reconstruction was quantified by the coding fraction, γ , which is a number between 0 and 1, with $\gamma = 1$ implying perfect reconstruction.

Partition experiment

The animal's head was electrically isolated from the trunk by a thin rubber partition so that stimuli could be presented to the head and trunk regions independently. The partition decreased the normal EOD amplitude in the head region because it partly blocked EOD current flow. To compensate for this a continuous unmodulated EOD mimic signal was delivered to the head region between a single electrode in the dorsal musculature and electrodes lateral to either side of the fish. This restored the EOD amplitude in the head region to values measured with the partition short-circuited. C_{low} and C_{high} values with the partition in place for local stimulation were not statistically different from those obtained without the partition ($P = 0.4359$ and $P = 0.95$ respectively, t -tests, $n = 15$).

Single electroreceptor afferents were recorded and stimulated with 4-Hz sinusoidal AMs to gauge the effectiveness of electrical isolation. Each afferent's responses to stimuli presented in the head and trunk regions were summarized as cycle histograms and quantified as vector strengths. The Rayleigh statistic²⁹ was also calculated to determine whether the cycle histogram showed statistically significant phase-locking to the AM. Receptor afferent receptive field positions ranged from 8 to 45 mm rostral or caudal to the partition. Stimulation of the receptor-containing region (1 mV cm^{-1} AM) resulted in a mean Rayleigh statistic of 758 ± 125 , showing significant phase-locking. Stimuli applied to the region not containing the receptor resulted in a mean Rayleigh statistic of 0.928 ± 0.321 . No receptor afferents recorded showed statistically significant phase-locking to the AM applied to the chamber not containing its receptive field.

The average distance between the partition and the CRF centre was 5.6 cm and was always greater than 4 cm. Using results reported previously⁶, we estimated the maximal CRF dimensions as follows. Average CRF centre and surround areas were summed, then doubled. The maximum distance from the centre of the receptive field to the CRF boundary was then estimated as the radius of a circle having this area (2.4 cm). This conservative estimate is lower than 4 cm. Along with the lack of pyramidal-cell responses to stimulation of the head chamber alone, this indicates that it is very unlikely that the CRF surround extends past the partition.

Received 20 January; accepted 17 March 2003; doi:10.1038/nature01590.

- Rieke, F., Bodnar, D. A. & Bialek, W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. Lond. B* **262**, 259–265 (1995).
- Machens, C. K. et al. Representation of acoustic communication signals by insect auditory neurons. *J. Neurosci.* **21**, 3215–3227 (2001).
- Nelson, M. E. & MacIver, M. A. Prey capture in the weakly electric fish *Apteronotus leptorhynchus*: sensory acquisition strategies and electrosensory consequences. *J. Exp. Biol.* **202**, 1195–1203 (1999).
- Zupanc, G. K. H. & Maler, L. Evoked chirping in the weakly electric fish *Apteronotus leptorhynchus*: a quantitative biophysical analysis. *Can. J. Zool.* **71**, 2301–2310 (1993).
- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Bastian, J., Chacron, M. J. & Maler, L. Receptive field organization determines pyramidal cell stimulus-encoding capability and spatial stimulus selectivity. *J. Neurosci.* **22**, 4577–4590 (2002).
- Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J. & Davis, J. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* **378**, 492–496 (1995).
- Vinje, W. & Gallant, J. L. Sparse Coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).
- Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
- Voss, R. F. & Clarke, J. '1/f noise' in music: music from 1/f noise. *J. Acoust. Soc. Am.* **63**, 258–263 (1978).
- Bastian, J. Electrolocation. I. How the electroreceptors of *Apteronotus albifrons* code for moving objects and other electrical stimuli. *J. Comp. Physiol. A* **144**, 465–479 (1981).
- Gabbiani, F., Metzner, W., Wessel, R. & Koch, C. From stimulus encoding to feature extraction in weakly electric fish. *Nature* **384**, 564–567 (1996).
- Maler, L., Sas, E. K. & Rogers, J. The cytology of the posterior lateral line lobe of high frequency weakly electric fish (*Gymnotoidei*): Dendritic differentiation and synaptic specificity in a simple cortex. *J. Comp. Neurol.* **195**, 87–139 (1981).
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. R. & Bialek, W. *Spikes: Exploring the Neural Code* (MIT, Cambridge, Massachusetts, 1996).
- Borst, A. & Theunissen, F. Information theory and neural coding. *Nature Neurosci.* **2**, 947–957 (1999).
- Mainen, Z. F. & Sejnowski, T. J. Reliability of spike timing in neocortical neurons. *Science* **268**, 1503–1506 (1995).
- de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R. & Bialek, W. Reproducibility and variability in neural spike trains. *Science* **275**, 1805–1808 (1997).
- Maler, L. & Mugnaini, E. Correlating gamma-aminobutyric acidergic circuits and sensory function in the electrosensory lateral line lobe of a gymnotiform fish. *J. Comp. Neurol.* **345**, 224–252 (1994).
- Berman, N. J. & Maler, L. Neural architecture of the electrosensory lateral line lobe: Adaptations for coincidence detection, a sensory searchlight and frequency-dependent adaptive filtering. *J. Exp. Biol.* **202**, 1243–1253 (1999).
- Crampton, W. G. R. Electric signal design and habitat preferences in a species rich assembly of gymnotiform fishes from the upper Amazon basin. *Anais Acad. Bras. Cienc.* **70**, 805–847 (1998).
- Zhang, H., Xu, J. & Feng, A. S. Effects of GABA-mediated inhibition on direction-dependent frequency tuning in the frog inferior colliculus. *J. Comp. Physiol.* **184**, 85–98 (1999).
- Macleod, K. & Laurent, G. Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assemblies. *Science* **274**, 976–979 (1996).
- Doiron, B., Chacron, M. J., Maler, L., Longtin, A. & Bastian, J. Inhibitory feedback required for network burst responses to communication but not prey stimuli. *Nature* **421**, 539–543 (2003).
- Carr, C. E., Maler, L. & Sas, E. Peripheral organization and central projections of the electrosensory organs in gymnotiform fish. *J. Comp. Neurol.* **211**, 139–153 (1982).
- Heiligenberg, W. & Dye, J. Labelling of electrosensory afferents in a gymnotid fish by intracellular injection of HRP: The mystery of multiple maps. *J. Comp. Physiol. A* **148**, 287–296 (1982).
- Metzner, W. & Heiligenberg, W. The coding of signals in the electric communication of the gymnotiform fish *Eigenmannia*: From electroreceptors to neurons in the torus semicircularis of the midbrain. *J. Comp. Physiol. A* **169**, 135–150 (1991).
- Metzner, W. & Juranek, J. A sensory brain map for each behavior? *Proc. Natl Acad. Sci. USA* **26**, 14798–14803 (1997).
- Heiligenberg, W. *Neural Nets in Electric Fish* (MIT, Cambridge, Massachusetts, 1991).
- Bastchelet, E. *Circular Statistics in Biology* (Academic, New York, 1981).
- Gabbiani, F. Coding of time varying signals in spike trains of linear and half-wave rectifying neurons. *Network Comput. Neural Sys.* **7**, 61–85 (1996).

Acknowledgements We thank A.-M. Oswald, J. Lewis and B. Lindner for their reading the manuscript. This research was supported by NSERC (M.J.C., B.D., A.L.), CIHR (L.M., A.L.) and NIH (J.B.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.C. (mchacron@physics.uottawa.ca).

The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria

Timothy D. Read^{*,†}, Scott N. Peterson^{*,‡}, Nicolas Tourasse^{§,¶}, Les W. Baillie^{*,†||}, Ian T. Paulsen^{*,¶}, Karen E. Nelson^{*}, Hervé Tettelin^{*}, Derrick E. Fouts^{*}, Jonathan A. Eisen^{*,¶}, Steven R. Gill^{*}, Erik K. Holtzapple^{*}, Ole Andreas Økstad^{§,¶}, Erlendur Helgason^{§,¶}, Jennifer Rillstone^{*}, Martin Wu^{*}, James F. Kolonay^{*}, Maureen J. Beanan^{*}, Robert J. Dodson^{*}, Lauren M. Brinkac^{*}, Michelle Gwinn^{*}, Robert T. DeBoy^{*}, Ramana Madpu^{*}, Sean C. Daugherty^{*}, A. Scott Durkin^{*}, Daniel H. Haft^{*}, William C. Nelson^{*}, Jeremy D. Peterson^{*}, Mihai Pop^{*}, Hoda M. Khouri^{*}, Diana Radune^{*}, Jonathan L. Benton^{*}, Yasmin Mahamoud^{*}, Lingxia Jiang^{*}, Ioana R. Hance^{*}, Janice F. Weidman^{*}, Kristi J. Berry^{*}, Roger D. Plaut^{*}, Alex M. Wolf^{*}, Kisha L. Watkins^{*}, William C. Nierman^{*}, Alyson Hazen^{*}, Robin Cline^{*}, Caroline Redmond[†], Joanne E. Thwaite[†], Owen White^{*}, Steven L. Salzberg^{*,¶}, Brendan Thomason[☆], Arthur M. Friedlander^{**}, Theresa M. Koehler^{††}, Philip C. Hanna[☆], Anne-Brit Kolstø^{§,¶} & Claire M. Fraser^{*,‡,§,¶}

^{*} The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

[†] Medical Biotechnology Center, University of Maryland Biotechnology Institute, Baltimore, Maryland 21201, USA

[‡] Department of Biochemistry, [§] Department of Microbiology and Tropical Medicine, [¶] Department of Pharmacology, The George Washington University, Eye Street, Washington DC 20052, USA

^{||} School of Pharmacy, University of Oslo N-0316, Oslo, Norway

^{||} Defence Science Technology Laboratory, Porton Down, Salisbury SP4 0JQ, UK

[¶] Johns Hopkins University, Charles and 34th Streets, Baltimore, Maryland 21218, USA

[☆] The Biotechnology Center of Oslo, Oslo N-0317, Norway

[☆] Department of Microbiology & Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

^{**} US Army Medical Research Institute for Infectious Diseases, Frederick, Maryland 21702, USA

^{††} Department of Microbiology and Molecular Genetics, University of Texas–Houston Health Science Center Medical School, University of Texas, Houston, Texas 77225, USA

Bacillus anthracis is an endospore-forming bacterium that causes inhalational anthrax¹. Key virulence genes are found on plasmids (extra-chromosomal, circular, double-stranded DNA molecules) pXO1 (ref. 2) and pXO2 (ref. 3). To identify additional genes that might contribute to virulence, we analysed the complete sequence of the chromosome of *B. anthracis* Ames (about 5.23 megabases). We found several chromosomally encoded proteins that may contribute to pathogenicity—including haemolysins, phospholipases and iron acquisition functions—and identified numerous surface proteins that might be important targets for vaccines and drugs. Almost all these putative chromosomal virulence and surface proteins have homologues in *Bacillus cereus*, highlighting the similarity of *B. anthracis* to near-neighbours that are not associated with anthrax⁴. By performing a comparative genome hybridization of 19 *B. cereus* and *Bacillus thuringiensis* strains against a *B. anthracis* DNA microarray, we confirmed the general similarity of chromosomal genes among this group of close relatives. However, we found that the gene sequences of pXO1 and pXO2 were more variable between strains, suggesting plasmid mobility in the group. The complete sequence of *B. anthracis* is a step towards a better understanding of anthrax pathogenesis.

B. anthracis has become notorious as a bioweapon because of its tough, environmentally resistant endospore and its ability to cause lethal inhalational anthrax. During the course of the disease,

endospores are taken up by alveolar macrophages where they germinate in the phagolysosomal compartment¹. Vegetative cells then escape from the macrophage, eventually infecting blood. Expression of the major plasmid-encoded virulence determinants, tripartite toxin and a poly-D-glutamic acid capsule, are essential for full pathogenicity¹. Sequencing the chromosome of *B. anthracis* was undertaken to help identify additional genes that might contribute to virulence either by encoding functions necessary for the survival and escape from the mammalian macrophage or by enhancing evasion of the immune system and the extent of damage caused by the bacterium to its animal host.

The *B. anthracis* Ames chromosome sequenced in this work (5,227,293 base pairs, bp) derives from an isolate taken from a dead cow in Texas (Methods). This sequence differs in only 11 confirmed single nucleotide polymorphisms⁵ (SNPs) from the 2001 Florida attack Ames isolate, verifying that the chromosome sequenced to completion is essentially identical to a virulent strain. The chromosome encodes 5,508 predicted protein-coding sequences (Table 1) with a pronounced bias for genes on the replication leading strand (Fig. 1), as has been seen in other low G + C Gram-positive replicons⁶. A feature shared with the chromosomes of other endospore-forming Gram-positive species of the genera *Bacillus* and *Clostridium*^{7,8} is the concentration of the ribosomal RNA, transfer RNA and ribosomal protein genes around the replication origin. This arrangement may maximize protein synthesis during early rounds of DNA replication after germination from the dormant endospore phase. The chromosome also contains at least four prophages (Supplementary Information) as well as two type I introns, one of which disrupts the *recA* gene⁹. Housekeeping functions such as DNA replication and fatty-acid metabolism are overwhelmingly partitioned to the chromosome, whereas the pXO1 and pXO2 plasmids have a greater proportion of transposons, genes involved in toxicity and genes without function assigned (Table 1).

Most *B. anthracis* Ames chromosomal proteins have homologues to proteins encoded on the draft genome sequence of *B. cereus* ATCC 10987 (T.D.R., unpublished results), a closely related strain (Figs 1 and 2). There are only 141 proteins in *B. anthracis* for which a putative functional assignment could not be made that do not have a match in the protein set of *B. cereus* ATCC 10987 sequence (BLASTP¹⁰ $E < 10^{-5}$). For the most part, these are encoded by genes of unknown function, are transposases or are present in phage regions. Almost all potential chromosomal virulence-enhancing genes have homologues in *B. cereus* ATCC 10987, suggesting that they are not specifically associated with the unique pathogenicity of *B. anthracis* but are part of the common arsenal of the *B. cereus* group of bacteria¹¹.

Table 1 Features of the *B. anthracis* Ames genome

Feature	Chromosome	pXO1*	pXO2*
Size (bp)	5,227,293	181,677	94,829
Number of genes	5,508	217	113
Replicon coding (%)	84.3	77.1	76.2
Average gene length (nt)	800	645	639
G+C content (%)	35.4	32.5	33.0
rRNA operons	11	0	0
tRNAs	95	0	0
sRNAs	3	2	0
Phage genes†	62	0	0
Transposon genes†	18	15	6
Disrupted reading frame‡	37	5	7
Genes with assigned function	2,762	65	38
Conserved hypothetical genes	1,212	22	19
Genes of unknown function	657	8	5
Hypothetical genes	877	122	51

*The complete, annotated pXO1 and pXO2 plasmids from an Ames strain isolated from the 2001 US bioterror attack were also resequenced recently at TIGR⁵ and have been included in this analysis.

†According to TIGR role categories.

‡Genes with indels or point mutations resulting in early termination, confirmed by resequencing.

The chromosome of *B. anthracis* Ames contains several homologues of genes known to be involved in *B. cereus* and *B. thuringiensis* pathogenesis. These include two channel-forming type III haemolysins (BA5701, BA2241) and a complex of three non-haemolytic enterotoxins (BA1887–1889). Several *B. anthracis* Ames proteins have sequence homology to proteins that contribute to the virulence of the Gram-positive pathogen *Listeria monocytogenes*¹². These include phosphatidyl-inositol-specific and phosphatidyl-choline-preferring phospholipase C (BA0677 and BA3891), internalin-like genes (BA1346 and BA1406), listeriolysin O (BA3355), sigma factor B (BA0992) and p60 extracellular protease (BA1952 and BA5474). The significance of these homologues may lie in the similarities in the pathways of intracellular survival and multiplication of *L. monocytogenes*, and the germination, survival and escape from macrophages by *B. anthracis*.

B. anthracis contains a gene encoding a homologue of the enhancin protein (BA3443), first described in baculoviruses that infect gypsy moths. Enhancin is a metalloprotease that boosts viral infectivity by degrading the mucin layer surrounding insect guts¹³. A homologue of *B. anthracis* enhancin is also found in the genome of *Yersinia pestis*, which survives in both mammals and insects¹⁴. *B. anthracis* also contains two homologues of *B. thuringiensis* immune inhibitor A metalloprotease (BA0672 and BA1295), which enhances virulence in insects through cleavage of bacteriocidal lectins¹¹. The presence of these genes may be evidence of an insect-infecting lifestyle in a recent ancestor.

Germination of the anthrax endospore is a key initial event in the *B. anthracis* infectious cycle. *B. anthracis* has seven (six chromosomal and one plasmid-borne) paralogues of the *gerA* family of tricystronic operons utilized by endospores to recognize the presence of specific small molecules to initiate the germination process¹⁵. Protection of DNA during dormancy and efficient DNA repair during germination are also believed to be important factors in endospore viability. *B. anthracis* has several homologues of the *Bacillus subtilis* small acid soluble DNA protection proteins, and the full complement of DNA repair proteins found in *B. subtilis*. *B. anthracis* also appears to have additional DNA repair capabilities focused on UV-induced DNA damage, with a unique deoxyribodipyrimidine photolyase gene (BA3180) and two, rather than one, UV dimer endonucleases. The photolyase is more closely related to enzymes from proteobacteria than those from other Gram-positive bacteria. The *B. anthracis* genome encodes several proteins that mitigate damage by free-oxygen radicals, including five catalases and three Fe-Mn superoxide dismutases. Other detoxification functions for which no obvious homologues could be found in *B. subtilis* include bromoperoxidase, thiolperoxidase, multiple thioredoxin proteins and a cytoplasmic Cu-Zn superoxide dismutase (SodC; BA5139). SodC has been shown to have a key role in the virulence of certain other intracellular bacteria, counteracting nitric oxide-mediated killing in the macrophage¹⁶.

The *B. anthracis* chromosome encodes a machinery for sporulation that is broadly similar to *B. subtilis*⁷. The proteins with the highest degree of sequence divergence between the species are endospore coat constituents and endospore polysaccharide biosynthesis components, suggesting altered composition of the outer surface. *B. subtilis* alternative sigma factors, which govern a cascade of events associated with cell development, are also generally conserved. One sigma factor missing in *B. anthracis* is *sigD*, which is essential for the expression of the flagellum operon¹⁷. However, *L. monocytogenes*, which is motile and carries a flagellum operon similar to *B. anthracis*, also lacks a *sigD* gene¹⁸.

Despite having numerous predicted secreted proteins encoded in its genome (Supplementary Information), *B. anthracis* is notable for paucity of extracellular protease activity under standard laboratory conditions¹⁹. One reason for this lack of protein secretion may lie in a mutation that affects regulation of gene expression: a nonsense mutation in the *plcR* positive regulator gene²⁰. In *B. thuringiensis*

and *B. cereus*, the *plcR* gene product is known to upregulate the production of numerous extracellular enzymes through binding at an upstream motif (TATGNAN₄TNCATA). Although the *B. anthracis plcR* homologue is truncated, there are 56 putative *plcR* binding motifs in the chromosome and 2 on pXO2. The extracellular protein genes downstream include phospholipases, enterotoxins and haemolysins (Table 2), and the *plcR* mutation has been shown to account for a dramatic reduction in lecithinase, protease and haemolysin production by *B. anthracis*¹⁹. However, it is possible that some PlcR-regulated gene products still contribute to virulence but are under alternative regulatory controls, as low-level expression of some of the genes in the PlcR regulon has been reported in *B. anthracis*¹⁹. There is another PlcR-family protein in the genome (BA0597) that might potentially function to complement expression under certain conditions.

The chromosome of *B. anthracis* contains three homologues of the sortase transpeptidase responsible for attachment of secreted proteins to peptidoglycans on the cell surface of Gram-positive

bacteria²¹, and also contains the *csaAB* genes for binding of proteins with S-layer homology (SLH) domains to polysaccharide. Using searches against models for the sortase attachment sites and SLH domains, 34 candidate surface proteins were identified (Supplementary Information). Two putative *B. anthracis* sortase-attached genes have internalin-like repeats¹¹. The potential role of most proteins with SLH domains on the surface of *B. anthracis* is unknown at present. However, these surface proteins may mediate as-yet-unknown interactions between *B. anthracis* and its external environment, and could be targets for vaccine and drug design.

The broad similarity in metabolic and transport genes of *B. anthracis* and *B. subtilis* (the model aerobic Gram-positive organism)⁷ suggests many common capabilities, yet there are a number of idiosyncrasies that may shed light on the ecology of *B. anthracis*. Compared to *B. subtilis*, *B. anthracis* appears to have an expanded capacity for amino-acid and peptide utilization. For instance, there are 17 ABC-type peptide binding proteins in

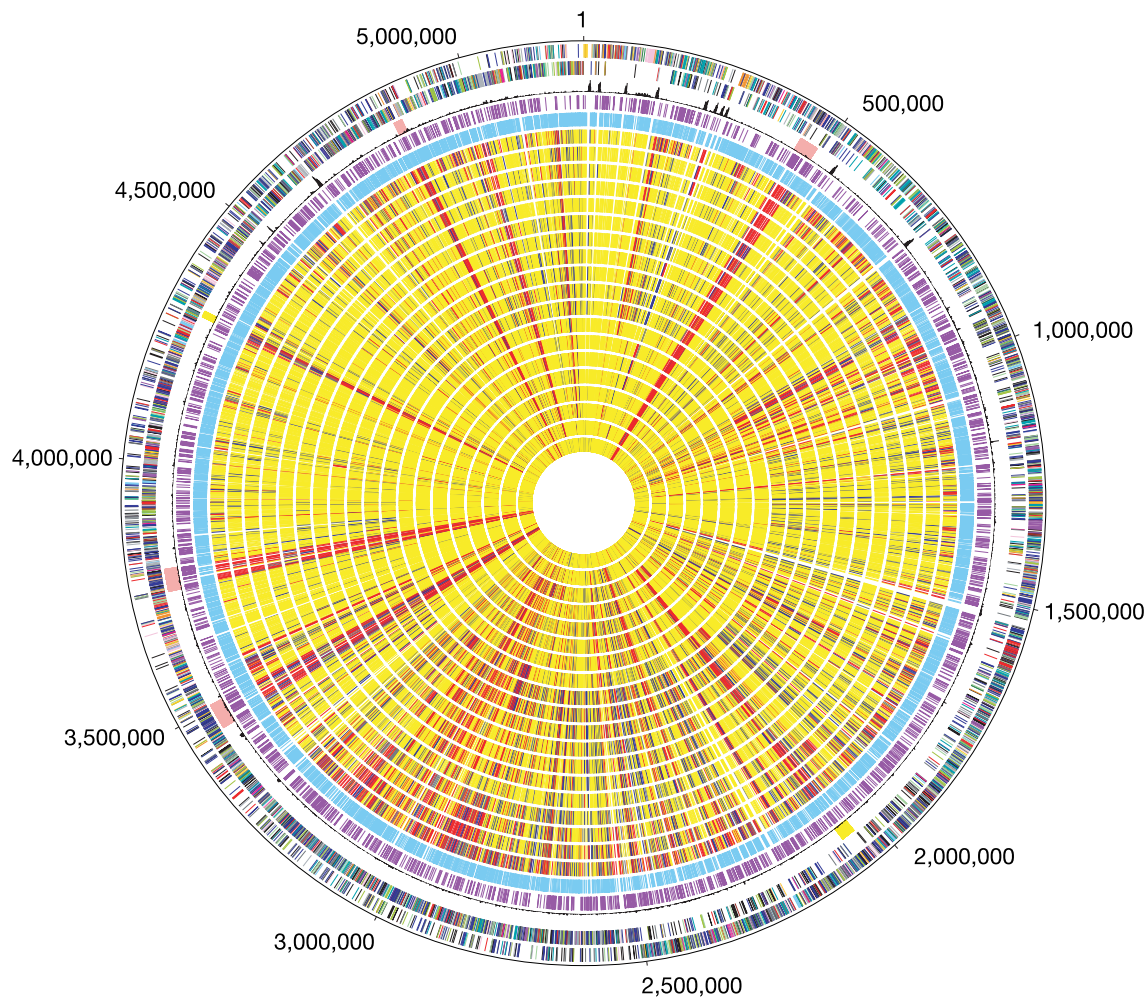


Figure 1 Circular representation of the *B. anthracis* chromosome and comparative genome hybridizations of *B. cereus* group strains. Outer circle, predicted coding regions on the plus strand colour-coded by role categories (see Supplementary Fig. 4). Circle 2, predicted coding regions on the minus strand colour-coded by role categories. Circle 3, atypical nucleotide composition curve. Salmon colour, phage regions; yellow, other unique regions located around positions 2.0 and 4.3 Mb (referred to as regions 5 and 6 in the text). Circle 4, genes not represented on the array. Circle 5, genes present on the array. Genes were classified into three groups: genes present in the query strain (shown

yellow), genes absent in the query strain (red), and diverged genes (blue). Missing data are in grey. *B. cereus* group strains are displayed following the phylogeny of Fig. 2 (circle number, strain number): 6, *B.c.* 874; 7, *B.c.* 535; 8, *B.c.* 612; 9, *B.w.* 1143; 10, *B.t.* 248; 11, *B.t.* 442; 12, *B.c.* 14579; 13, *B.t.* 775; 14, *B.c.* 259; 15, *B.t.* 1031; 16, *B.t.* 251; 17, *B.c.* 607; 18, *B.c.* ATCC 10987; 19, *B.c.* 812; 20, *B.c.* 819; 21, *B.c.* 831; 22, *B.t.* 840; 23, *B.c.* 1123; 24, *B.c.* 816. Here we use *B.c.*, *B.t.* and *B.w.* to indicate *B. cereus*, *B. thuringiensis* and *Bacillus weihenstephanensis*, respectively.

B. anthracis compared with four in *B. subtilis*; and there are nine homologues of the BrnQ branched chain amino-acid transporter in *B. anthracis* and only two in *B. subtilis*. *B. anthracis* also has an expanded number of secreted proteases and peptidases relative to *B. subtilis* and a number of amino-acid utilization genes not found to date in other *Bacillus* genomes such as homogentisate dioxygenase (BA0242), involved in tyrosine degradation. Emphasizing the potential importance of peptides and amino acids for *B. anthracis* metabolism, there are six LysE/Rht amino-acid efflux systems compared with two in *B. subtilis*. These systems prevent accumulation of amino acids to bacteriostatic concentrations during growth on peptides²². *B. anthracis* may therefore be adapted for life in a protein-rich environment, such as decaying animal matter.

B. anthracis appears to have a reduced capacity for sugar utilization relative to *B. subtilis*. It lacks catabolic pathways for mannose, arabinose and rhamnose, and has reduced numbers of phosphotransferase systems and other types of sugar transporters. *B. anthracis* possesses genes for the cleavage of extracellular chitin and chitosan, and the utilization of N-acetylglucosamine constituents of these polymers. This may reflect some type of association with insects analogous to *B. thuringiensis*, or with polymers derived from plant or fungal material. *B. anthracis* contains a complete operon for polyester biosynthesis, which may function as an alternative energy storage compound for the organism. *B. anthracis*

also has a multisubunit NADH hydrogenase not described before in Gram-positive bacteria.

B. anthracis possesses an expanded array of iron-acquisition genes compared to *B. subtilis* that may be important for iron scavenging in a mammalian host. These include 15 ABC uptake systems for iron siderophores or chelates, as well as two clusters of genes for the biosynthesis of siderophores. Two genes involved in synthesis of an aerobactin-like siderophore are not found in *B. subtilis* or the *B. cereus* ATCC 10987 sequence (BA1981, BA1982). Like *B. subtilis* and other soil bacteria, *B. anthracis* encodes a broad swathe of predicted drug efflux pumps, and a variety of other antibiotic-resistance genes are also present. However, it is unknown whether these contribute to resistance in a clinical setting²³.

We designed a *B. anthracis* DNA microarray on the basis of identifiable genes present at the conclusion of random phase sequencing. The microarray was used to compare *B. anthracis* to 19 members of the *B. cereus* group by comparative genome hybridization (CGH) (Fig. 1). Strains examined by CGH possessed 66–92% of their chromosomal genes in common with *B. anthracis*. Genes unique to *B. anthracis* in particular, and the *B. cereus* group in general, appear to be over-represented in the 2.0-Mb chromosomal region (coordinates 1,500,000 to 3,500,000 in Fig. 1) surrounding the presumed terminus of replication. Genome plasticity around the replication terminus has been seen in other comparisons of bacterial genomes²⁴. Six smaller regions of the *B. anthracis* genome appeared to be absent from nearly all other *B. cereus* group strains tested (Fig. 1). Regions one to four correspond to the *B. anthracis* prophages (Supplementary Information), and region five centres on an IS110 family insertion element. Only region six does not bear obvious relationship to mobile elements. The magnitude of genomic variability based on CGH experiments comparing the *B. anthracis* Ames microarray and *B. cereus* group strains (Fig. 1) is 25–100 times greater than in similar experiments involving comparison of *B. anthracis* Ames to other *B. anthracis* strains (T. Blank and S.N.P., unpublished results). This reflects the very limited molecular diversity of the *B. anthracis* species²⁵.

Hybridization experiments indicate the presence of pXO1 homologues in half of the 19 strains examined (Supplementary Information), consistent with what has been shown in other studies²⁶. Few genes from the pXO1 pathogenicity island, pXO1-96 to pXO1-127 (ref. 2), appeared to be present in the 19 *B. cereus* group strains. The toxin genes, central to anthrax aetiology, are found only in *B. anthracis* and not in any of the 19 *B. cereus* group strains sampled. In sharp contrast to pXO1, there were few pXO2 genes hybridizing with genomic DNA from the 19 *B. cereus* group bacteria.

Ratios obtained by CGH for chromosomal genes were used to infer the phylogenetic relationships among the *B. cereus* strains, producing three clusters (I, IIa and IIb; Fig. 2). The groupings were compatible with the phylogeny reconstructed using multilocus enzyme electrophoresis⁴. By extrapolation of the results from that study to the microarray-based phylogeny, *B. anthracis* would emerge within cluster IIa (arrow in Fig. 2). The presence of genes with pXO1 sequence identity in various branches covering all three clusters of the *B. cereus* group tree (Fig. 2), and the distribution of pXO1-like genes in the *B. cereus* group independent of the chromosomal relatedness among the strains (Supplementary Information), provides further evidence for mobility of pXO1 genes within the *B. cereus* group. Plasmid transfer within the *B. cereus* group is well established²⁷, and there are numerous mobility genes on pXO1². Despite the evidence for genomic variability in the *B. cereus* group, the *B. anthracis* chromosome and virulence plasmids display little localized variation in G + C content and dinucleotide composition (generally associated with horizontally acquired genes from distantly related donors), suggesting that most genes are native to the *B. cereus* group.

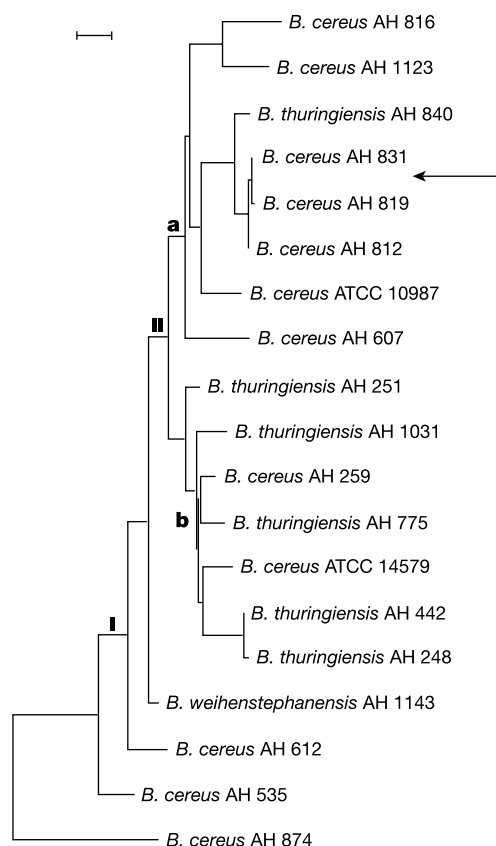


Figure 2 Phylogenetic relationships among 19 *B. cereus*/*B. thuringiensis* strains inferred from CGH results for 3,601 chromosomal *B. anthracis* genes. The tree was built by applying the neighbour-joining algorithm to a pairwise distance matrix of percentages of differences between the presence/absence patterns of all strains (diverged genes not taken into account). Similar trees were obtained using the maximum-parsimony method. The scale bar represents 2% divergence. The arrow indicates the position where *B. anthracis* would emerge by extrapolation from multilocus enzyme electrophoresis analysis⁴.

The *B. anthracis* chromosome sequence portrays a soil-dwelling organism, possessing numerous potential virulence genes, which has possibly a preference for protein-rich environments. This is consistent with the evolution of *B. anthracis* from a *B. cereus* ancestor through acquisition of key plasmid-encoded toxin, capsule and regulatory loci. CGH data presented here demonstrate variability in plasmid gene content among the group as compared to

chromosomal genes. Other major differences between *B. anthracis* and *B. cereus* may have been effected through altered gene expression rather than loss or gain of genes. Although both species contain genes associated with secreted proteases, haemolysins, extracellular chitinases¹¹, motility, tyrosine degradation and penicillin resistance²³, *B. anthracis* and *B. cereus* phenotypes differ with respect to the function of these genes. These changes in expression

Table 2 Putative PlcR-regulated proteins

Gene	Description	Motif-to-gene distance (nt)*	Signal peptide†
BA0401	Tellurium resistance protein	109	No
BA0400	Tellurium resistance protein		No
BA0399	Tellurium resistance protein, putative		No
BA0575	Methyl-accepting chemotaxis protein	101	Yes
BA0969	Hypothetical protein	55	No
BA0975	HD domain protein	143	No
BA0976	Hypothetical protein	219	No
BA0977	Hypothetical protein	157	Yes
BA1086	Sugar binding transcriptional regulator LacI family	73	No
BA1085	Acetyltransferase, GNAT family		No
BA1424	Histidyl-tRNA synthetase, putative	721	No
BA1470	Membrane protein, putative	747	No
BA1692	Conserved hypothetical protein	171	No
BA1888	Enterotoxin	520	Yes
BA1887	Enterotoxin		Yes
BA1889	Enterotoxin (point mutation)		Yes
BA2147	ScdA protein	16	No
BA2148	Hypothetical protein	268	No
BA2499	Hypothetical protein	123	Yes
BA2730	Neutral protease	114	Yes
BA3355	Thiol-activated cytolysin	248	Yes
BA3356	Membrane protein, putative	159	No
BA3370	Ribonuclease	269	Yes
BA3491	Conserved hypothetical protein	144	No
BA3635	Spore germination protein	36	Yes
BA3634	Spore germination protein		Yes
BA3633	Spore germination protein		No
BA3891	1-Phosphatidylinositol phosphodiesterase	106	Yes
BA3890	Hypothetical protein		No
BA3892	Serine protease, subtilase family	149	Yes
BA3893	Cell wall hydrolase, putative	146	Yes
BA4745	ABC transporter, ATP-binding protein	286	No
BA4744	Membrane protein, putative		Yes
BA4743	Rrf2 protein family protein, putative		No
BA4746	Acid phosphatase	58	Yes
BA4949	Metallo- β -lactamase family protein	84	No
BA5055	Conserved domain protein	194	No
BA5190	Acetyltransferase GNAT family	93	No
BA5191	NAD(P)H dehydrogenase, quinone family	518	No
BA5231	Hypothetical protein	14	No
BA5230	Hypothetical protein		No
BA5243	CAAX amino terminal protease family protein	139	Yes
BA5595	Transcriptional regulator plcR-related, putative, authentic point mutation	102	No
BA5594	plcR-associated protein		Yes
BA5596	Hypothetical protein	313	No
BA5605	Hypothetical protein	113	No
BA5606	Aminopeptidase, putative	113	Yes
BA5701	Channel protein, haemolysin III family	6	Yes
BA5700	UDP-galactose-4-epimerase		No
BA5702	Conserved hypothetical protein	179	No

*Distance from the 5' end of the motif to the first nucleotide in the first gene of the operon. Genes in the same block are in the same predicted operon (that is, same transcription orientation and with no intervening rho-dependent terminator).

†See Supplementary Methods.

may reflect recent adaptations following acquisition of the pathogenicity island that contains the lethal toxin loci on pXO1. The *atxA* regulatory gene in this region controls toxin gene expression but is incompatible with the chromosomal regulator *plcR*, found in *B. cereus*¹⁹. The worldwide, near-clonal spread of the organism²⁵ suggests that expression of the toxin and capsule genes confers an advantage to *B. anthracis* that outweighs changes in the chromosomal gene expression. Findings from this genome sequence analysis raise further questions about the biology of *B. anthracis*; for instance, what are the roles of putative 'virulence' genes in close relatives of *B. anthracis* that do not cause anthrax, and do they actually contribute to virulence in *B. anthracis*? □

Methods

Genome sequencing and analysis of *B. anthracis* Ames (pXO1⁺ pXO2⁺)

B. anthracis Ames was cured of plasmid pXO1 by incubation at 43 °C and pXO2 subsequently cured by novobiocin treatment (Supplementary Methods). As previously described, the chromosome was sequenced using two DNA preparations. For the first (Porton1)⁵, 2–3 kilobase (kb) and 4–7 kb random insert libraries in plasmid-derived vectors were constructed and end-sequenced following the standard strategy for TIGR microbial shotgun projects⁶, achieving success rates of 74% and 64% and average high-quality read lengths of 559 nucleotides (nt) and 586 nt, respectively. For the Porton2 preparation³, libraries of 2–3 kb and 6–8 kb were constructed with success rates of 89% and 85% and average high-quality read lengths of 609 nt and 645 nt. The completed chromosome sequence consisted of 73,806 and 6,052 reads from the Porton1 small and large insert libraries, and 3,532 and 32,430 from the Porton2 small and large insert libraries—achieving an average of 13-fold sequence coverage per base. After assembly, gaps between contigs were closed by editing, walking library clones, and linking assemblies by polymerase chain reaction (PCR). The Glimmer gene finder²⁸ was modified by enhancing its model of noncoding sequences. This improved its ability to exclude short open reading frames (ORFs), and substantially reduced the number of predicted small hypothetical proteins. Annotation was as described for a previous project⁶. BLASTP¹⁰ was used for comparisons of the protein sets of *B. anthracis*, *B. cereus* ATCC 10987 (T.D.R., unpublished results; <http://www.tigr.org/tdb/ufmg/>) and other complete bacterial genomes (<http://www.tigr.org/cmr2/> and <http://genolist.pasteur.fr/Subtilist/>). A predicted probability score of less than 10^{−5} was used as a standard cut-off to define a likely match.

DNA microarray preparation and analysis

Amplicons representing 79 of 217 and 41 of 122 genes from pXO1 and pXO2 respectively, and 3,601 of 5,753 chromosomal genes as predicted by Glimmer²⁸ (see Supplementary Methods) were arrayed onto glass microscope slides (Telechem Inc.). Redundant genes were generally represented once or a few times on the array. Genomic DNA was labelled with Cy3 and Cy5 according to J. DeRisi (http://www.microarrays.org/Pdfs/GenomicDNAlabel_B.pdf), except that genomic DNA was not digested or sheared before labelling. Arrays were scanned with a GenePix 4000B scanner (Axon Inc.). Hybridization signals were quantified using TIGR SPOTFINDER (software available at <http://www.tigr.org/softlab>). Hybridization experiments were competitive using probes derived from *B. anthracis* Ames (reference) and a *B. cereus* group (query) strain. Normalized signal intensities were used to generate relative hybridization ratios (query/reference). Data representing weak signal were removed. The ratios from a maximum of six data points (duplicate spots, hybridizations performed in triplicate) were placed in three bins: <0.1, gene is absent in query strain; 0.1–0.3, present but diverged in query strain; and >0.3, gene is present in the query strain. A majority rule was applied to the data for binning such that more than 50% of ratios were in agreement as to assignment and that at least two data points were used (exceeded in 99% of the cases). In cases where less than two data points existed, the gene was treated as data missing.

The criteria for the numerical ranges of our bins were established in two ways. First, we determined the presence or absence of sequences homologous to 3,601 *B. anthracis* genes in the sequence of *B. cereus* ATCC 14579 (Integrated Genomics Inc.; <http://www.integratedgenomics.com/>) using BLASTN¹⁰, and compared that to the assignments inferred from hybridization ratios. A threshold of 0.1 was found to be suitable for classifying a gene as absent (that is, agreement between sequence and CGH data in 99% of the cases), while a cut-off value of 0.3 was conservative for gene presence (agreement in 92% of the cases). Second, we used a set of 65 genes conserved in 26 bacterial genomes, NCBI COG database (<http://www.ncbi.nlm.nih.gov/COG/>). Genes judged as present in query strains using our selected cut-offs correctly binned data in 1,225 out of 1,235 total calls. There was a tendency for underprediction of plasmid homologues by CGH, when compared to results from the sequence analysis. Two possible explanations for this are variability in plasmid copy number in *B. cereus* strains relative to *B. anthracis* and/or that the average divergence of plasmid genes is greater than chromosomal genes.

Other techniques and analysis

PCR amplification for microarray spotting, pulsed field gel electrophoresis and Southern blotting are described in Supplementary Information, as is the phylogenetic analysis of chromosomal data.

Received 14 August 2002; accepted 28 March 2003; doi:10.1038/nature01586.

- Dixon, T. C., Meselson, M., Guillemin, J. & Hanna, P. C. Anthrax. *N. Engl. J. Med.* **341**, 815–826 (1999).
- Okinaka, R. T. *et al.* Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J. Bacteriol.* **181**, 6509–6515 (1999).
- Okinaka, R. T. *et al.* Sequence, assembly and analysis of pXO1 and pXO2. *J. Appl. Microbiol.* **87**, 261–262 (1999).
- Helgason, E. *et al.* *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* **66**, 2627–2630 (2000).
- Read, T. D. *et al.* Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* **296**, 2028–2033 (2002).
- Tettelin, H. *et al.* Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506 (2001).
- Kunst, F. *et al.* The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256 (1997).
- Nolling, J. *et al.* Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**, 4823–4838 (2001).
- Ko, M., Choi, H. & Park, C. Group 1 self-splicing intron in the *recA* gene of *Bacillus anthracis*. *J. Bacteriol.* **184**, 3917–3922 (2002).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Guttmann, D. M. & Ellar, D. J. Phenotypic and genotypic comparisons of 23 strains from the *Bacillus cereus* complex for a selection of known and putative *B. thuringiensis* virulence factors. *FEMS Microbiol. Lett.* **188**, 7–13 (2000).
- Cossart, P. Molecular and cellular basis of the infection by *Listeria monocytogenes*: an overview. *Int. J. Med. Microbiol.* **291**, 401–409 (2002).
- Lepore, L. S., Roelvink, P. R. & Granados, R. R. Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. *J. Invertebr. Pathol.* **68**, 131–140 (1996).
- Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001).
- McCann, K. P., Robinson, C., Sammons, R. L., Smith, D. A. & Corfe, B. M. Alanine germination receptors of *Bacillus subtilis*. *Lett. Appl. Microbiol.* **23**, 290–294 (1996).
- De Groote, M. A. *et al.* Periplasmic superoxide dismutase protects *Salmonella* from products of phagocyte NADPH-oxidase and nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **94**, 13997–14001 (1997).
- West, J. T., Estacio, W. & Marquez-Magana, L. Relative roles of the *fla/che* P(A), P(D-3), and P(sigD) promoters in regulating motility and *sigD* expression in *Bacillus subtilis*. *J. Bacteriol.* **182**, 4841–4848 (2000).
- Glaser, P. *et al.* Comparative genomics of *Listeria* species. *Science* **294**, 849–852 (2001).
- Mignot, T. *et al.* The incompatibility between the *PlcR* and *AtxA*-controlled regulons may have selected a nonsense mutation in *Bacillus anthracis*. *Mol. Microbiol.* **42**, 1189–1198 (2001).
- Agaisse, H., Gominet, M., Okstad, O. A., Kolsto, A. B. & Lereclus, D. *PlcR* is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol. Microbiol.* **32**, 1043–1053 (1999).
- Pallen, M. J., Lam, A. C., Antonio, M. & Dunbar, K. An embarrassment of sortases—a richness of substrates? *Trends Microbiol.* **9**, 97–102 (2001).
- Bellmann, A. *et al.* Expression control and specificity of the basic amino acid exporter LysE of *Corynebacterium glutamicum*. *Microbiology* **147**, 1765–1774 (2001).
- Turnbull, P. C. Definitive identification of *Bacillus anthracis*—a review. *J. Appl. Microbiol.* **87**, 237–240 (1999).
- Suyama, M. & Bork, P. Evolution of prokaryotic gene order: genome rearrangements in closely related species. *Trends Genet.* **17**, 10–13 (2001).
- Keim, P. *et al.* Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J. Bacteriol.* **182**, 2928–2936 (2000).
- Pannucci, J., Okinaka, R. T., Sabin, R. & Kuske, C. R. *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. *J. Bacteriol.* **184**, 134–141 (2002).
- Thorne, C. B. *Bacillus subtilis* and Other Gram-positive Bacteria 113–124 (American Society for Microbiology, Washington DC, 1993).
- Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **26**, 544–548 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We acknowledge the contributions of P. Turnbull, E. Saile, Y. Chen, J. Hunter-Cevera, N. McKinney, S. Cendrowski, M. Weiner, A. Fouet, A. Harrison, S. Leppla, M. Mock, C. Moran, G. Myers, G. Patra, J. Ravel, E. Reilly and T. Torok. The *B. anthracis* chromosome sequence was supported by funding from the Office of Naval Research (ONR), National Institutes of Allergy and Infectious Disease (NIAID), the Department of Energy and the UK Defence Science Technology Laboratory. Comparative genome hybridization experiments were supported by the ONR. A-B.K., N.T., O.A.O. and E.H. were supported by the Norwegian Research Council. The sequencing of *B. cereus* ATCC 10987 was supported by the NIAID under the Pathogen Functional Genomics Resource Center contract, and the ONR.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.D.R. (anthrax@tigr.org). The *B. anthracis* genome sequence has been deposited at GenBank under accession number AE016879; the microarray data have been deposited at Gene Expression Omnibus (GEO) under accession number GSE341. The *B. cereus* 10987 unfinished genome sequence (GenBank accession number NC_003909) is available at <http://www.tigr.org/tdb/ufmg>.

Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*

Natalia Ivanova*, Alexei Sorokin†, Iain Anderson*, Nathalie Galleron†, Benjamin Candelon†, Vinayak Kapatral*, Anamitra Bhattacharyya*, Gary Reznik‡, Natalia Mikhailova*, Alla Lapidus*, Lien Chu*, Michael Mazur*, Eugene Goltsman*, Niels Larsen*, Mark D'Souza*, Theresa Walunas*, Yuri Grechkin*, Gordon Pusch*, Robert Haselkorn*, Michael Fonstein*, S. Dusko Ehrlich†, Ross Overbeek* & Nikos Kyrpides*

* Integrated Genomics, Chicago, Illinois 60612, USA

† G n tique Microbienne, Centre de Recherche de Jouy-en-Josas, Institut National de la Recherche Agronomique, 78352 Jouy en Josas cedex, France

‡ Life Sciences Operation, Illinois Institute of Technology Research Institute, Chicago, Illinois 60616, USA

§ Deceased

Bacillus cereus is an opportunistic pathogen causing food poisoning manifested by diarrhoeal or emetic syndromes¹. It is closely related to the animal and human pathogen *Bacillus anthracis* and the insect pathogen *Bacillus thuringiensis*, the former being used as a biological weapon and the latter as a pesticide. *B. anthracis* and *B. thuringiensis* are readily distinguished from *B. cereus* by the presence of plasmid-borne specific toxins (*B. anthracis* and *B. thuringiensis*) and capsule (*B. anthracis*). But phylogenetic studies based on the analysis of chromosomal genes bring controversial results, and it is unclear whether *B. cereus*, *B. anthracis* and *B. thuringiensis* are varieties of the same species² or different species^{3,4}. Here we report the sequencing and analysis of the type strain *B. cereus* ATCC 14579. The complete genome sequence of *B. cereus* ATCC 14579 together with the gapped genome of *B. anthracis* A2012⁵ enables us to perform comparative analysis, and hence to identify the genes that are conserved between *B. cereus* and *B. anthracis*, and the genes that are unique for each species. We use the former to clarify the phylogeny of the *cereus* group, and the latter to determine plasmid-independent species-specific markers.

The general features of the *B. cereus* ATCC 14579 genome are listed in Table 1. The region between 0.8 and 1.8 megabases (Mb) has G + C content close to the average for the chromosome (35.3%); in the region between 3.7 and 0.8 Mb the G + C content is higher than the average value, and in the region between 1.8 and 3.7 Mb it is lower than the average value (Fig. 1, circle 3). These regions are bordered with putative prophages (Fig. 1, circles 6 and 7), which could be indicative of the origin of the *B. cereus* ATCC

14579 chromosome as a result of phage-mediated recombination between the chromosomes of closely related bacteria.

A 15.1-kilobase (kb) contig with G + C content of 38%, which could not be joined to any other by the multiplex long accurate (MLA) polymerase chain reaction (PCR) procedure or assembled into a separate circular structure by long-range (LR) PCR, corresponds to the linear plasmid, detected earlier in total DNA preparations from this bacterium⁶. One of the coding sequences (CDSs) in this contig is homologous to a type B DNA polymerase typically found in *Bacillus subtilis* phage Φ 29 and several linear mitochondrial plasmids. The contig also contains a CDS with similarity to endolysin. We therefore suggest that it corresponds to a linear plasmid or prophage, which we designated pBClin15. The termini of the plasmid could be protected by covalently closed hairpin telomeres or by a covalently bound protein, as in the phage Φ 29. However, pBClin15 carries no CDS with homology to the phage terminal protein. The polymerase gene of pBClin15 is more closely related to those of mitochondrial linear plasmids than to polymerases of *B. subtilis* phages.

Phylogenetic analysis of the *cereus* group²⁻⁴ shows that *B. cereus* ATCC 14579 and *B. anthracis* strains are not particularly close, and the ecological niches occupied by these bacteria are also very different⁷. Nevertheless 4,505 CDSs in *B. cereus* ATCC 14579 have 80–100% identity to their homologues in *B. anthracis* A2012. The average identity for this group of genes is 92.1%. Some general statistics of the genome comparison are also provided in Table 1 (rows 10–14). Potential orthologues identified as bidirectional best hits represent approximately 75% of the CDSs for each of the genomes. 85% of these potential orthologues have conserved neighbourhood, suggesting extensive short-range synteny (Table 1).

The large core set of genes (75–80%) conserved between *B. cereus* ATCC 14579 and *B. anthracis* A2012 could have been inherited from a common ancestor. Analysis of the metabolic potential encoded by the core set contradicts the hypothesis of the *cereus* group common ancestor being a soil bacterium. The characteristic feature of soil bacteria, such as *Streptomyces* spp. or *B. subtilis*, is the multiplicity of carbohydrate catabolic pathways reflecting the variety of carbohydrates in the soil, where plant-derived material is the main source of nutrients. Whereas a total of 41 genes for degradation of carbohydrate polymers were identified in the *B. subtilis* genome, only 14 and 15 CDSs coding for polysaccharide degradation enzymes are present in *B. cereus* and *B. anthracis*, respectively, and the spectrum of polysaccharides that can be degraded by these bacteria is limited to glycogen and starch, chitin and chitosan (Supplementary Fig. 1). In contrast, the abundance of proteolytic enzymes, the multiplicity of peptide and amino-acid transporters and the variety of amino-acid degradation pathways (Supplementary Table 1) indicates that proteins, peptides and amino acids may

Table 1 General features of *B. cereus* ATCC 14579 and *B. anthracis* A2012 genomes

Row	Property	<i>B. cereus</i>	<i>B. anthracis</i>
1	DNA contigs (plasmids)	2 (1)	3* (2)
2	DNA sequenced (bp)	5,426,909	5,370,060
3	Coding sequence (bp)	4,559,996 (84.0%)	4,357,132 (84.1%)
4	G + C content (%)	35.3%	35.1%
5	RNA operons	13	?
6	CDSs, total	5,366 (100%)	5,842 (100%)
7	CDSs with assigned function	3,839 (71.5%)	4,173 (71.4%)
8	Conserved hypothetical CDSs	1,481 (26.8%)	1,563 (26.8%)
9	CDSs with no similarity	142 (2.7%)	109 (1.9%)
10	CDSs with bidirectional best hits†	4,302 (80.2%)	4,302 (73.6%)
11	CDSs in shared protein families‡	4,690 (87.4%)	4,969 (85%)
12	<i>B. cereus</i> CDSs not found in <i>B. anthracis</i>	860 (16.0%)	
13	<i>B. anthracis</i> CDSs not found in <i>B. cereus</i>		874 (15%)
14	CDSs with bidirectional best hits† in conserved chromosomal clusters	3,652 (84.9%)	3,652 (84.9%)

* Refers to the genomic scaffold of the *B. anthracis* chromosome that has 417 gaps.

† Identified as pairs of close bidirectional best hits²⁰.

‡ Cut-off score used is 10⁻⁵.

be a preferred nutrient source for *B. cereus* and *B. anthracis*. A total of 51 and 48 protease-encoding CDSs were identified in *B. cereus* and *B. anthracis*, respectively, compared to only 30 in *B. subtilis*. While three of the *B. subtilis* extracellular proteases are absent from

B. cereus (Epr, Bpr and AprX), other proteases, which are found in one or two copies in *B. subtilis*, are represented as large families in *B. cereus* and *B. anthracis* (Supplementary Table 1). Several observations suggest that the insect intestine could have been the natural

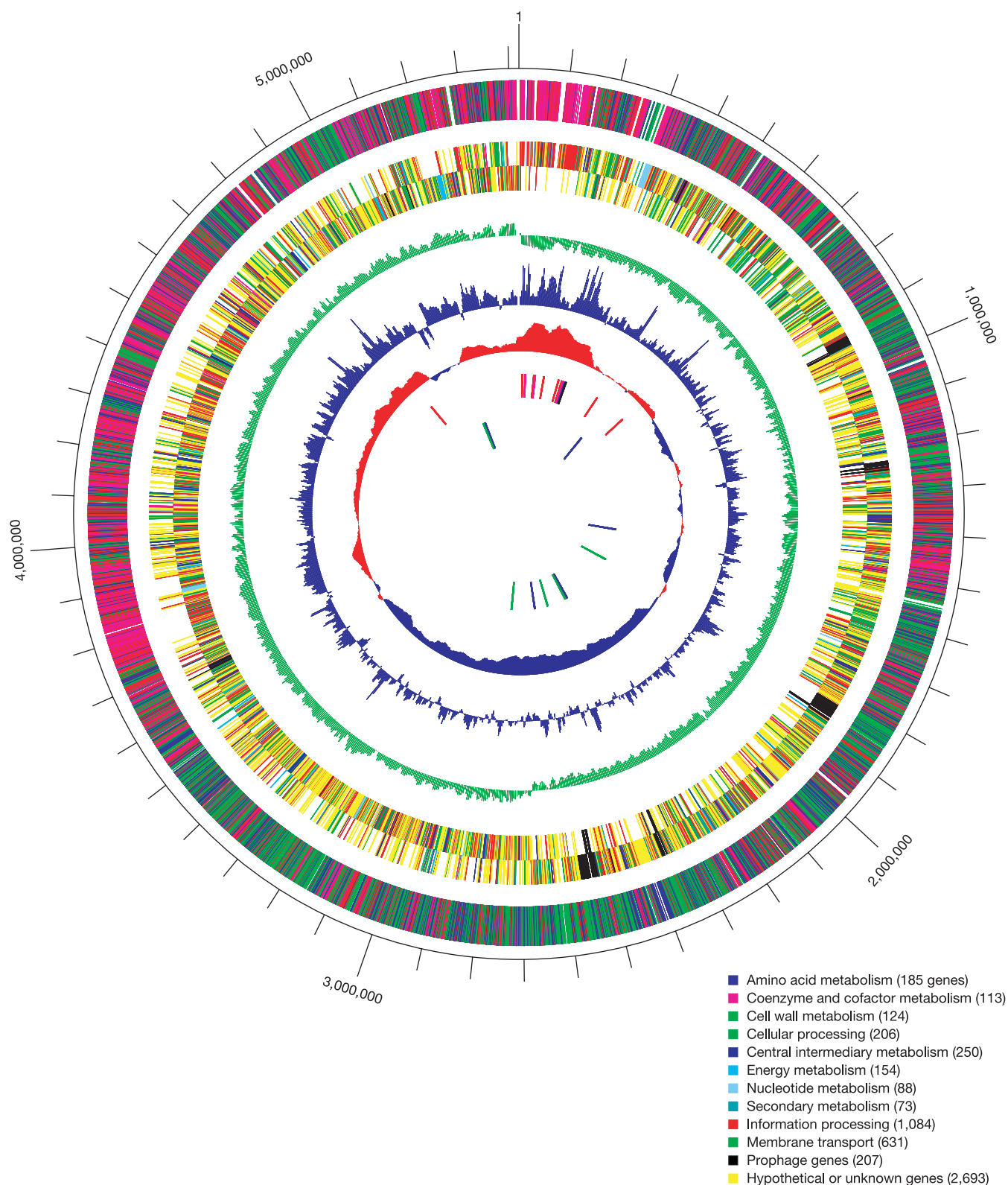


Figure 1 Genome map of *B. cereus* ATCC 14579. From the inside: green and blue bars show positions of two insertion sequence (IS) elements, red-to-black bars show positions of *rmn* operons. Circle 1, G + C content over 200-kb window with 5-kb step; red and blue denote respectively G + C content higher and lower than average. Circles 2 and 3,

G + C and (C – G)/(C + G) content over 20-kb window with 5-kb step. Circles 4 and 5, CDSs on the – and + strands, colour reflects functional category (see key). Circle 6, homology to *B. subtilis* CDSs using FASTA. Green, 0–30% identity; blue, 30–45%; red, 45–65%; pink, 65–100%.

habitat for the common ancestor of the *cereus* group⁸. The peritrophic membrane of insect guts consists of chitin and protein components. While the chitinolytic enzymes enable *B. cereus* and *B. anthracis* to degrade the chitin component, homologues of the zinc metalloprotease enhancin⁹ of entomopathogenic viruses (BC3384 and BA_3939) could provide the ability to cleave the invertebrate intestinal mucin, which is the major protein component of the insect peritrophic membrane. A potential PlcR-binding site was found upstream of the enhancin homologue in *B. cereus*.

The core set of genes conserved between *B. cereus* and *B. anthracis* includes numerous factors for invasion, establishment and propagation of bacteria within the host. While the presence of such genes in *B. anthracis* is not surprising, finding of pathogenicity-related genes in *B. cereus* ATCC 14579 was somewhat unexpected. Genes encoding all but two toxins ever identified in *B. cereus* clinical isolates were found in the reputedly non-pathogenic *B. cereus* ATCC 14579 (Supplementary Table 2). Both *B. anthracis* and *B. cereus* implement several mechanisms of protection against the host defence system. Three homologues of the immune inhibitor A protein (InhA), which selectively cleaves insect antibacterial peptides¹⁰, were found in *B. cereus* ATCC 14579, and two homologues of InhA are present in *B. anthracis* A2012. A homologue of staphylococcal MprF protein¹¹ found in *B. cereus* (BC1465) and *B. anthracis* (BA_2009) could further enhance resistance to antibacterial peptides by decreasing the negative charge of the bacterial surface via aminoacylation of anionic phospholipids with lysine. The counterparts of the S-layer proteins from *B. anthracis* and *B. thuringiensis* were not found in the *B. cereus* ATCC 14579 genome, although several CDSs with SLH domains were identified. This observation is in agreement with the results of Kotiranta *et al.*¹², who demonstrated

that, unlike many clinical isolates, the strain ATCC 14579 is devoid of an S-layer. The presence of CDSs encoding potential pathogenicity factors in *B. cereus*, *B. anthracis* and *B. thuringiensis*^{13,14} is consistent with the *cereus* group ancestor being an opportunistic insect pathogen rather than a benign soil bacterium.

The repair of ultraviolet (UV)-induced DNA damage could have been of critical importance to the *cereus* group ancestor, as there are numerous mechanisms to repair UV-induced lesions. Direct reversal of UV-damage is mediated by two photolyases: one is spore-specific (*splB*-type) and the other is of the *phrB*-type. *B. cereus* utilizes a bacterial *uvrABC* repair system for UV-dimer excision, and is also capable of transcription-coupled repair (BC0058). In addition, *B. cereus* has two CDSs (BC0260, BC5347) homologous to the C-terminal endonuclease domain of the UvdE UV-repair endonuclease from *Neurospora crassa* and Uve1 from *Schizosaccharomyces pombe*.

The pleiotropic regulator PlcR was previously identified as one of the principal regulators of *B. cereus* virulence genes^{13–15}. When no mismatches were allowed, a total of 55 possible PlcR-binding sites were found, of which 26 coincide with the promoter regions of genes and 24 were found in the upstream regions of potential operons, bringing the number of genes that could be controlled directly by PlcR to more than 100 (Supplementary Table 3). In addition to PlcR (BC5350), at least four transcriptional regulators (BC1715, BC2410, BC2770 and BC3194) belong to the potential PlcR regulon (Fig. 2), suggesting that other CDSs could be regulated by PlcR indirectly, which is in accord with recent proteomics data¹⁶. The presence of three PlcR paralogues (BC0988, BC1158 and BC2443) in the genome of *B. cereus* ATCC 14579 could make the PlcR regulatory network even more complex. Though none of the

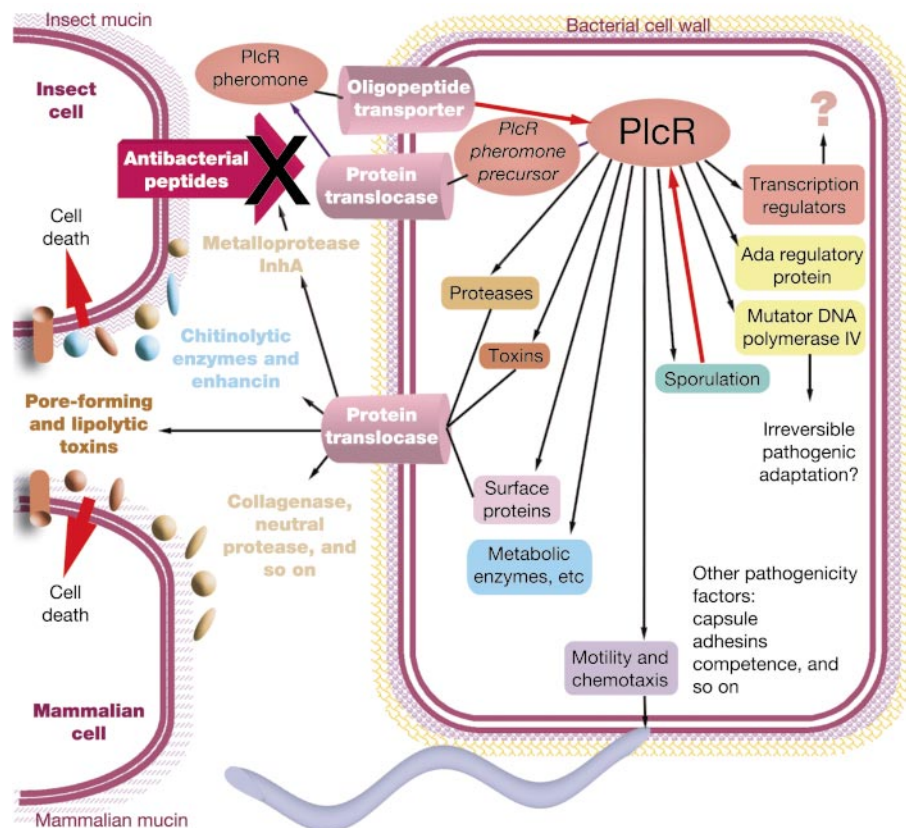


Figure 2 Schematic representation of the potential PlcR regulon based on the presence of putative PlcR-binding sites upstream of the genes and operons. Genes with putative PlcR-binding sites in the upstream regions were classified into several categories, including toxins, proteases, surface proteins, motility and chemotaxis genes, antibiotic efflux proteins,

and so on. Activation of the PlcR regulon results in secretion of numerous proteins with cytotoxic activities towards mammalian and insect cells. Expression of the *plcR* gene is activated by an unknown mechanism, which includes oligopeptide permease-dependent uptake of a PapR peptide encoded by *CDS2* located downstream from the PlcR gene.

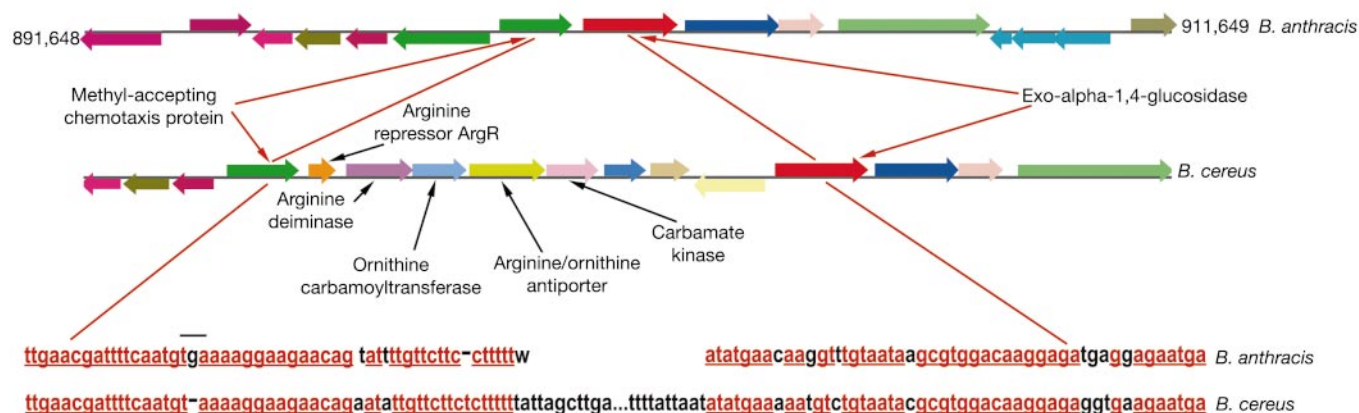


Figure 3 Deletion of the arginine deiminase operon from *B. anthracis*. The top portion of the figure shows CDSs within the equivalent genomic regions from *B. anthracis* and *B. cereus*. The orthologues of CDSs (shown as arrows of the same colour) flanking the

arginine deiminase operon in *B. cereus* are adjacent to each other in the *B. anthracis* genome. Below are the DNA sequences of *B. anthracis* and *B. cereus* flanking the deletion site. Conserved nucleotides are shown red.

PlcR paralogues appears to be controlled directly by PlcR (Supplementary Table 3), it is possible that they are activated by the PapR peptide, as was demonstrated for PlcR¹⁷.

The histidine protein kinase (BC3528) homologous to the *B. subtilis* sporulation kinase KinB appears to be an important member of the PlcR regulatory network. It is known that the phosphorylated form of the transition state regulator Spo0A acts as a repressor of PlcR¹⁸, so upregulation of BC3528 by PlcR would provide a feedback mechanism that controls PlcR expression. An orthologue of protein kinase BC3528 is absent from *B. anthracis* A2012, which could contribute to the incompatibility of PlcR and AtxA regulons in *B. anthracis*¹⁹. Another potential member of the PlcR regulon, the mutator DNA polymerase IV (BC4142), could provide irreversible pathogenic adaptation of *B. cereus* and *B. anthracis*. DNA polymerase IV belongs to the DinB/UmuCD/Rad30/Rev1 superfamily of error-prone DNA polymerases that have been shown to induce adaptive mutability in bacteria under stressful conditions²⁰. We suggest that during host colonization, polymerase IV-dependent mutagenesis could result in adaptive point mutations that enhance survival of *B. cereus* and related bacteria, and potentially affect their pathogenicity.

Approximately 15% of the CDSs in either genome show no similarity with the other genome (Table 1). Prophage proteins and transposases account for approximately 140 unique CDSs in each organism. We detected six prophages in the genome of ATCC 14579 that were not previously characterized in the *B. cereus* group. The prophages are not inserted within the transfer RNA operons, and they do not carry any of the known pathogenicity factors of *B. cereus*.

Some of the unique CDSs found in *B. cereus* ATCC 14579 and *B. anthracis* A2012 appear to be specific for a certain group of strains rather than species-specific. In *B. cereus* ATCC 14579, a chromosomal cluster that could code for capsular polysaccharide biosynthesis was found. It covers more than 20 kb (BC5279–BC5263), and contains genes for glycosyltransferases and flippase-type translocase, as well as polysaccharide polymerization machinery including chain length regulator and its regulatory protein-tyrosine kinase and phosphatase. A cluster specific for *B. anthracis* A2012 (BA_0356 to BA_0371), which contains several genes for glycosyltransferases and an ABC transporter similar to the teichoic acid exporter *tagGH* of *B. subtilis*, could code for biosynthesis of a teichuronic acid or a secondary cell wall polymer that serves as an anchor for S-layer proteins.

B. cereus ATCC 14579 has an extensive repertoire of restriction-modification systems, which includes type I, II and 5-methylcytosine-specific systems. The *B. cereus* type I system comprises restriction, methylation and specificity subunits. The type II system has

restriction and methylation domains; the type II R domains are weakly similar to McrB-type and *Lactococcus lactis* LlaI restriction systems. *B. anthracis* lacks the type I and II restriction-modification systems, but, unlike *B. cereus*, it has a CDS weakly similar to the 5-methylcytosine-specific Mrr endonuclease.

A chromosomal cluster (BC5090–BC5083) potentially coding for biosynthesis of a novel peptide antibiotic was found in *B. cereus* ATCC 14579, but not in *B. anthracis*. The cluster includes the precursor peptide and putative modification proteins, including a homologue of the subtilin biosynthesis protein SpaB (BC5084) that catalyses dehydration of serine and threonine. However, no homologue of the thioether-forming SpaC protein was found, suggesting an unusual structure for this peptide antibiotic.

Other CDSs identified as unique for *B. cereus* ATCC 14579 and *B. anthracis* A2012 could be used as plasmid-independent species-specific markers. A seven-gene chromosomal cluster for inositol degradation is present in *B. anthracis*, but not in *B. cereus*. Compared to the *B. subtilis* *iol* operon, two genes, *iolH* and *iolE*, are missing from the *B. anthracis* cluster, which probably makes the *B. anthracis* operon non-functional, as all *iol* genes except *iolS* are essential for inositol utilization in *B. subtilis*²¹. In *B. cereus*, a chromosomal cluster encoding the enzymes from the arginine deiminase pathway (BC0406–BC0409) was identified. This pathway enables *Streptococcus pyogenes* to survive acidic conditions in the presence of arginine owing to release of ammonium²², and it may play a similar role in *B. cereus*. In *B. anthracis*, the entire arginine deiminase cluster appears to be deleted (Fig. 3), though the neighbouring CDSs are conserved. Ammonium inhibits receptor-mediated internalization of the lethal toxin²³, therefore ammonium production by arginine deiminase could be disadvantageous for *B. anthracis*. Deletion of the arginine deiminase operon in *B. anthracis* could be a case of disposal of genes detrimental for the pathogenic lifestyle, similar to that of *Escherichia coli* lysine decarboxylase²⁴.

The availability of both complete and gapped genome sequence data from bacteria belonging to the *cereus* group provides a basis for whole-genome-based phylogenetic analysis, with a view to exploring the genetic diversity within the *cereus* group, and inferring the nature and origin of the species. These sequence data should also facilitate the identification of genes that are crucial for host colonization and bacterial propagation during *B. cereus* infection; such findings would have implications for understanding the biology of *B. anthracis*. □

Methods

B. cereus strain ATCC 14579 was obtained from ATCC, and used to construct libraries in plasmids (2–3-kb inserts in pGEM3) and cosmids (30–35-kb inserts using Loris 6). The

strain 6A5, which is considered to be the same as ATCC 14579, was obtained from BGSC, and used in Génétique Microbienne, INRA, to perform LR PCR for the final stages of the project. High-molecular-mass genomic DNA was isolated from *B. cereus* following standard protocols. The DNA was either used for LR PCR, or sheared, size-fractionated, and used to construct libraries. Whole-genome shotgun sequencing was performed on about 30,000 plasmids and 3,000 cosmids using Applied Biosystems 3700 DNA sequencers (Perkin-Elmer). In the first phase of sequencing, the genome was assembled using Phred-Phrap-Consed into 547 contigs longer than 2,000 base pairs (bp), the longest being 50,451 bp. Gaps were closed by primer walking over cosmid clone inserts (3,000 reactions, resulting in 128 contigs longer than 2,000 bp, the longest being 273,559 bp) and by sequencing of LR PCR products using the finishing strategy based on MLA PCR²⁵ (2,320 reactions). Finally, the genome was assembled into two contigs representing the circular chromosome and the linear plasmid with an average of 6 × coverage. The presence of a linear plasmid was confirmed by Southern hybridization. The statistical distribution of random readings between plasmid and chromosomal DNA indicates that the plasmid is present in one copy per chromosome (130 reads for the 15-kb plasmid compared to 47,315 reads for the 5,412-kb chromosome in one of the assembly versions).

The number of *rrn* operons was determined by LR PCR and Southern hybridization (not shown), and each operon was sequenced by primer walking. Thirteen *rrn* operons were found (Fig. 1), this number being higher than 5 to 12 operons reported earlier for strains of *B. cereus* or *Bacillus weihenstephanensis*^{26,27}. The sequences of 16S, 23S and 5S ribosomal RNAs are very similar in all operons, the number of differences being up to 2, 7 and 4 bases, respectively. No 16S rRNA gene containing the psychrotolerance (ability to grow at low temperature) signature was detected. The corresponding *rrn* operons of *B. anthracis* A2012 were omitted from the deposited sequence, so no comparison could be made at this point.

Genome sequence data for *B. anthracis* A2012 (NC_003995) was obtained from NCBI. Genes were identified by combination of Critica and a CDS-calling program developed at Integrated Genomics (IG). Nine additional CDSs on the chromosome of *B. anthracis* were identified, compared to the annotation provided by NCBI, by searching intergenic DNA sequences against the IG genome database. The genomes of *B. cereus* and *B. anthracis* were subjected to a round of automatic annotation followed by extensive manual curation within the ERGO bioinformatics suite²⁸. Comparative analysis of *B. cereus* and *B. anthracis* genomes was carried out with the WorkBench algorithm (Integrated Genomics), as described for three strains of *Xylella fastidiosa*²⁹. As the sequence of *B. anthracis* chromosome is incomplete, the CDSs identified as unique for *B. cereus* were subjected to further analysis, including analysis of the chromosomal context and search for homologues in the unfinished *B. anthracis* and *B. cereus* genomes in the TIGR database (<http://www.tigr.org>) and in the gapped genome of *B. thuringiensis israelensis*³⁰. ERGO tools for pattern matching were used to identify the genes potentially controlled by the virulence regulator PlcR^{13–15} based on the finding of the consensus sequence TATGnAnnnnTnCAT¹⁵ in the upstream region of the genes. The results of analysis, including annotations and partial metabolic reconstructions of *B. anthracis* A2012 and *B. cereus* ATCC 14579 can be found in the limited version of ERGO at <http://www.ergo-light.com>.

Received 13 August 2002; accepted 21 March 2003; doi:10.1038/nature01582.

- Kotiranta, A., Lounatmaa, K. & Haapsalo, M. Epidemiology and pathogenesis of *Bacillus cereus* infections. *Microbes Infect.* **2**, 189–198 (2000).
- Helgason, E. *et al.* *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* **66**, 2627–2630 (2000).
- Ticknor, L. O. *et al.* Fluorescent amplified fragment length polymorphism analysis of norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. *Appl. Environ. Microbiol.* **67**, 4863–4973 (2001).
- Vilas-Boas, G., Sanchis, V., Lereclus, D., Lemos, M. V. F. & Bourguet, D. Genetic differentiation between sympatric populations of *Bacillus cereus* and *Bacillus thuringiensis*. *Appl. Environ. Microbiol.* **68**, 1414–1424 (2002).
- Read, T. D. *et al.* Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* **296**, 2028–2033 (2002).
- Carlson, C. R., Grønstad, A. & Kolsto, A. B. Physical map of the genomes of three *Bacillus cereus* strains. *J. Bacteriol.* **174**, 3750–3756 (1992).
- Turnbull, P. C. B. Definitive identification of *Bacillus anthracis*—a review. *J. Appl. Microbiol.* **87**, 237–240 (1999).
- Margulis, L. *et al.* The Arthromitus stage of *Bacillus cereus*: intestinal symbionts of animals. *Proc. Natl Acad. Sci. USA* **95**, 1236–1241 (1998).
- Wang, P. & Granados, R. R. An intestinal mucin is the target substrate for a baculovirus enhancin. *Proc. Natl Acad. Sci. USA* **94**, 6977–6982 (1997).
- Fedhila, S., Nel, P. & Lereclus, D. The InhA2 metalloprotease of *Bacillus thuringiensis* strain 407 is required for pathogenicity in insects infected via the oral route. *J. Bacteriol.* **184**, 3296–3304 (2002).
- Peschel, A. *et al.* *Staphylococcus aureus* resistance to human defensins and evasion of neutrophil killing via the novel virulence factor MprF is based on modification of membrane lipids with L-lysine. *J. Exp. Med.* **193**, 1067–1076 (2001).
- Kotiranta, A. *et al.* Surface structure, hydrophobicity, phagocytosis, and adherence to matrix proteins of *Bacillus cereus* cells with and without the crystalline surface protein layer. *Infect. Immun.* **66**, 4895–4902 (1998).
- Agaisse, H., Gominet, M., Økstad, O. A., Kolsto, A. B. & Lereclus, D. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol. Microbiol.* **32**, 1043–1053 (1999).
- Salamitou, S. *et al.* The plcR regulon is involved in the opportunistic properties of *Bacillus thuringiensis* and *Bacillus cereus* in mice and insects. *Microbiology* **146**, 2825–2832 (2000).
- Økstad, O. A. *et al.* Sequence analysis of three *Bacillus cereus* loci carrying PlcR-regulated genes encoding degradative enzymes and enterotoxin. *Microbiology* **145**, 3129–3138 (1999).
- Gohar, M. *et al.* Two-dimensional electrophoresis analysis of the extracellular proteome of *Bacillus cereus* reveals the importance of the PlcR regulon. *Proteomics* **2**, 784–791 (2002).
- Slamti, L. & Lereclus, D. A cell-cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. *EMBO J.* **21**, 4550–4559 (2002).
- Lereclus, D., Agaisse, H., Grandvalet, C., Salamitou, S. & Gominet, M. Regulation of toxin and

- virulence gene transcription in *Bacillus thuringiensis*. *Int. J. Med. Microbiol.* **290**, 295–299 (2000).
- Mignot, T. *et al.* The incompatibility between the PlcR- and AtxA-controlled regulons may have selected a nonsense mutation in *Bacillus anthracis*. *Mol. Microbiol.* **42**, 1189–1198 (2001).
- McKenzie, G. J., Lee, P. L., Lombardo, M. J., Hastings, P. J. & Rosenberg, S. M. SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification. *Mol. Cell* **7**, 571–579 (2001).
- Yoshida, K.-I., Aoyama, D., Ishio, I., Shibayama, T. & Fujita, Y. Organization and transcription of the *myo*-inositol operon, *iol*, of *Bacillus subtilis*. *J. Bacteriol.* **179**, 4591–4598 (1997).
- Degnan, B. A. *et al.* Characterization of an isogenic mutant of *Streptococcus pyogenes* Manfredo lacking the ability to make streptococcal acid glycoprotein. *Infect. Immun.* **68**, 2441–2448 (2000).
- Gordon, V. M., Leppla, S. H. & Hewlett, E. L. Inhibitors of receptor-mediated endocytosis block the entry of *Bacillus anthracis* adenylate cyclase toxin but not that of *Bordetella pertussis* adenylate cyclase toxin. *Infect. Immun.* **56**, 1066–1069 (1988).
- Maurelli, A. T., Fernandez, R. E., Bloch, C. A., Rode, C. K. & Fasano, A. “Black holes” and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 3943–3948 (1998).
- Sorokin, A. *et al.* A new approach using multiplex long accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and sequencing. *Genome Res.* **6**, 448–453 (1996).
- Johansen, T., Carlson, C. R. & Kolsto, A. B. Variable number of rRNA operons in *Bacillus cereus* strains. *FEMS Microbiol. Lett.* **136**, 325–328 (1996).
- Lechner, S. *et al.* *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. *Int. J. Syst. Bacteriol.* **48**, 1373–1382 (1998).
- Kapatral, V. *et al.* Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586. *J. Bacteriol.* **184**, 2005–2018 (2002).
- Bhattacharyya, A. *et al.* Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. *Proc. Natl Acad. Sci. USA* **99**, 12403–12408 (2002).
- Anderson, I. *et al.* Genome of *Bacillus thuringiensis* subsp. *israelensis* and comparative genomics of the *Bacillus cereus* group. *J. Bacteriol.* (submitted).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This Letter is dedicated to the memory of our dear colleagues M. Mazur and C. Anagnostopoulos. This work was supported by a DARPA STTR grant to Integrated Genomics Inc.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.I. (ivanova@integratedgenomics.com). The complete genome sequence of *B. cereus* ATCC 14579 has been deposited in GenBank at accession numbers AE016877 (chromosome) and AE016878 (plasmid).

An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene

Shinji Hirotsune[†]*, Noriyuki Yoshida^{*}, Amy Chen[‡], Lisa Garrett[‡], Fumihiro Sugiyama[§], Satoru Takahashi[§], Ken-ichi Yagami[§], Anthony Wynshaw-Boris[‡] & Atsushi Yoshiki[¶]

* Division of Neuroscience, Research Center for Genomic Medicine, Saitama Medical School Yamane 1397-1, Hidaka City, Saitama 350-1241, Japan
† PRESTO, Japan Science and Technology Corporation, Hon-cho 4-1-8, Kawaguchi, Saitama, Japan

‡ Genetic Disease Research Branch, National Human Genome Research Institute, NIH, Bldg 49, Room 4C80 49, Convent Drive, Bethesda, Maryland 20892, USA
§ Institute of Basic Medical Sciences and Laboratory Animal Resource Center, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8575, Japan
¶ Departments of Pediatrics and Medicine, UCSD Cancer Center, University of California, San Diego School of Medicine, 9500 Gilman Drive, Mailstop 0627, La Jolla, California 92093-0627, USA

¶ Experimental Animal Division, Department of Biological Systems, BioResource Center, RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

A pseudogene is a gene copy that does not produce a functional, full-length protein¹. The human genome is estimated to contain up to 20,000 pseudogenes^{2,3}. Although much effort has been devoted to understanding the function of pseudogenes, their biological roles remain largely unknown. Here we report the

role of an expressed pseudogene—regulation of messenger-RNA stability—in a transgene-insertion mouse mutant exhibiting polycystic kidneys and bone deformity. The transgene was integrated into the vicinity of the expressing pseudogene of *Makorin1*, called *Makorin1-p1*. This insertion reduced transcription of *Makorin1-p1*, resulting in destabilization of *Makorin1* mRNA *in trans* by way of a *cis*-acting RNA decay element within the 5' region of *Makorin1* that is homologous between *Makorin1* and *Makorin1-p1*. Either *Makorin1* or *Makorin1-p1* transgenes could rescue these phenotypes. Our findings demonstrate a specific regulatory role of an expressed pseudogene, and point to the functional significance of non-coding RNAs.

In the course of making a transgenic mouse carrying the *Drosophila sex-lethal* gene, we found interesting phenotypes in heterozygotes of one line of transgenic mice. Approximately 80% of heterozygotes died within two days of birth. Survivors of this critical period suffered from failure to thrive, and exhibited severe bone deformities (Fig. 1a). We measured bone mineral density (BMD) of fibia from independent heterozygotes by single-energy X-ray absorption measurements⁴. All heterozygotes exhibited reduced mineralization of 50–70% compared to wild-type littermates (Fig. 1b). The cortex of tibial bone was remarkably attenuated, and its trabeculae were delicate strands with disorganization (Fig. 1c). The bone showed hypercellularity with mosaic lines and reduction in osteoid matrix. Chondrocyte columnation appeared normal, with a relative increase in proliferating chondrocytes. These histological and macroscopic features are similar to those found in osteogenesis imperfecta. Mutant heterozygotes also developed progressive renal polycystic dilation (Fig. 1d), with microscopic liver cysts (data not shown). Finally, mutant mice displayed a skin defect in the embryo (Fig. 1e). The epithelial cover on the eyes was not completed, resulting in an opening in later stage embryos. Although we made several transgenic lines, and confirmed expression of the transgene in all lines, only one particular line exhibited these pleiotropic phenotypes. Thus, we concluded that these phenotypes could probably be attributed to gene inactivation by transgene insertion, rather than a function of the transgene.

We used relatively mildly affected heterozygous transgenic males to mate with wild-type females to maintain the mutant line. In

order to analyse homozygotes, we mated heterozygous mutants to each other. Surprisingly, there was no significant difference of phenotypic severity between homozygous and heterozygous mutants, suggesting that the mutated gene was regulated by imprinting. To test this hypothesis, we first analysed F₁s from matings between heterozygous transgenic males and wild-type females. In this case, 99 out of 108 (92%) heterozygotes exhibited the kidney and bone phenotypes, including perinatal lethality (Fig. 1f left). By contrast, only 3 out of 96 (3%) heterozygotes from matings between wild-type males and heterozygous transgenic females exhibited the phenotypes (Fig. 1f right). Thus, the parental origin of mutant allele was a critical factor for phenotypic severity, suggesting an imprinted transmission pattern. To confirm the reversibility of imprinting, we mated the unaffected male heterozygotes (derived from matings between wild-type males and mutant females) with wild-type females. These heterozygous offspring displayed a high proportion (98 out of 109 heterozygotes, 90%) of the mutant phenotypes (Fig. 1f right). Therefore, we concluded that the gene mutated by transgene insertion was an imprinted gene.

To identify the gene responsible for the phenotypes, we constructed a cosmid library from the genomic DNA of a mutant mouse. We screened approximately 8×10^5 independent colonies using the transgene as a probe, and isolated three independent clones that carried the transgene. We further screened a wild-type bacterial artificial chromosome (BAC) library using genomic regions flanking the transgene. The chromosome location of the transgene was mapped in radiation hybrids, demonstrating that the transgene was tightly linked to *D5Mit237*, a location confirmed by fluorescence *in situ* hybridization (FISH) analysis using the BAC clone (data not shown). The genome structure of the transgene locus was characterized (Fig. 2a). One copy of each of two transgenes was integrated into the genome in a head-to-head fashion, accompanied by a 7-kilobase (kb) deletion. There were no obvious chromosome rearrangements beyond the transgene insertion site. Next, we analysed the BAC clone spanning the transgene locus to identify the responsible gene. First, we sequenced the BAC clone. The sequence information was used to search expressed-sequence-tag (EST) databases, and was subjected to exon prediction programs. In addition, several candidates were subjected to northern

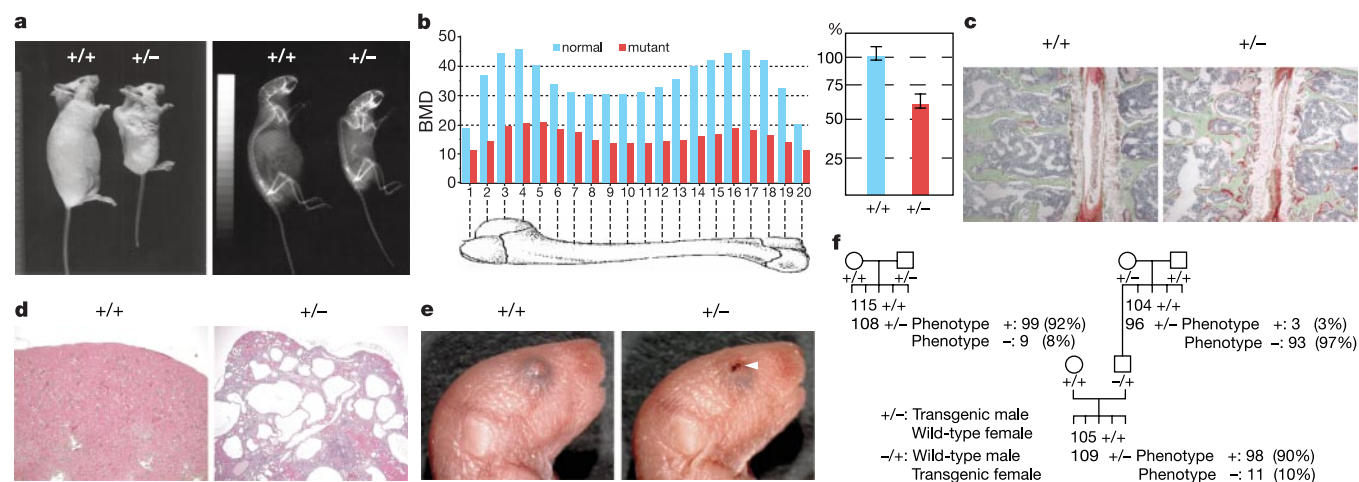


Figure 1 Analysis of mutant phenotypes and genetic study. **a**, Macroscopic findings (left) and soft X-ray examination (right) are shown for wild-type (+/+) or transgenic (+/-) three-month-old mice. **b**, Left, tibial bone mineral density (BMD) was measured in a mutant and its littermate by single-energy X-ray absorption measurements. Right, relative density was analysed in 5 pairs of the +/- and +/+ littermates between 2 and 4 months of age. **c**, Histological examination of thoracic vertebrae (Villanueva's Goldner method) from three-month-old mutant (+/-) and control (+/+) mice. **d**, Histological examination of kidney (haematoxylin-eosin) from three-month-old mutant (+/-) and control (+/+) mice. **e**, At P0, all heterozygotes (+/-) exhibited loss of eye-closure seen in wild types (+/+). **f**, Left, +/- males were mated with +/+ females, and 99 of 108 (92%) +/- offspring displayed the phenotypes. Right, +/- males were mated with +/- females, but only 3 of 96 (3%) +/- offspring displayed the phenotypes. Unaffected heterozygous males (-/+) from this mating were crossed with +/+ females. Transgenic offspring of this mating displayed a high frequency of the phenotypes (90%).

mice. Severe cystic dilatation was observed in the heterozygote. **e**, At P0, all heterozygotes (+/-) exhibited loss of eye-closure seen in wild types (+/+). **f**, Left, +/- males were mated with +/+ females, and 99 of 108 (92%) +/- offspring displayed the phenotypes. Right, +/- males were mated with +/- females, but only 3 of 96 (3%) +/- offspring displayed the phenotypes. Unaffected heterozygous males (-/+) from this mating were crossed with +/+ females. Transgenic offspring of this mating displayed a high frequency of the phenotypes (90%).

blot analysis from mouse RNA and Southern blot analysis using DNAs from several other species to find conserved regions (zoo blot). Using this information, we found three transcriptional units located in the vicinity of the transgene locus (Fig. 2a). Each candidate was tested for imprinted expression. There were no significant differences of transcript levels of either transcript 1 or *deoxycytidine kinase* (*DCK*) among the F₁s of the different mating combination, so we excluded these two genes as candidates (Fig. 2b). We focused on the analysis of transcript 2, which was located at the position closest to the transgene. Although transcript 2 (*Makorin1-p1*: Genebank accession number, AF494488) exhibited the typical structure of a truncated pseudogene of *Makorin1*, *Makorin1-p1* was robustly transcribed. We isolated a complementary DNA of *Makorin1-p1* from a random primed cDNA library of mouse embryo. Sequence analysis indicated that *Makorin1-p1* encoded an approximately 700-base-pair (bp) fragment which overlapped with the 5' region of *Makorin1* mapped to murine chromosome 6 (ref. 5) (Fig. 2c). The sequence of *Makorin1-p1* also

indicated that numerous nucleotide changes were accumulated, and the open reading frame was fragmented, suggesting that *Makorin1-p1* could not be translated into a protein. Although the precise 3' end of *Makorin1-p1* was not clear, the size of the isolated cDNA was highly consistent with the size of *Makorin1-p1* RNA by northern analysis, suggesting that the isolated cDNA was almost full-length. We further analysed the *Makorin1-p1* transcription unit using labelled sense or anti-sense oligonucleotides. A sense transcript was produced, but there was no detectable anti-sense transcript (data not shown).

Next, we examined the transcription level of the *Makorin1* gene on chromosome 6 and the *Makorin1-p1* pseudogene on chromosome 5. *Makorin1* was originally identified as a gene with strong homology with *ZNF127*, which is located in the critical region of the Prader-Willi syndrome^{5,6}. There are two kinds of transcripts of *Makorin1* generated by alternative splicing. Although the transcripts of *Makorin1* contain different 3' untranslated regions (UTRs), the protein coding regions are identical. The shorter

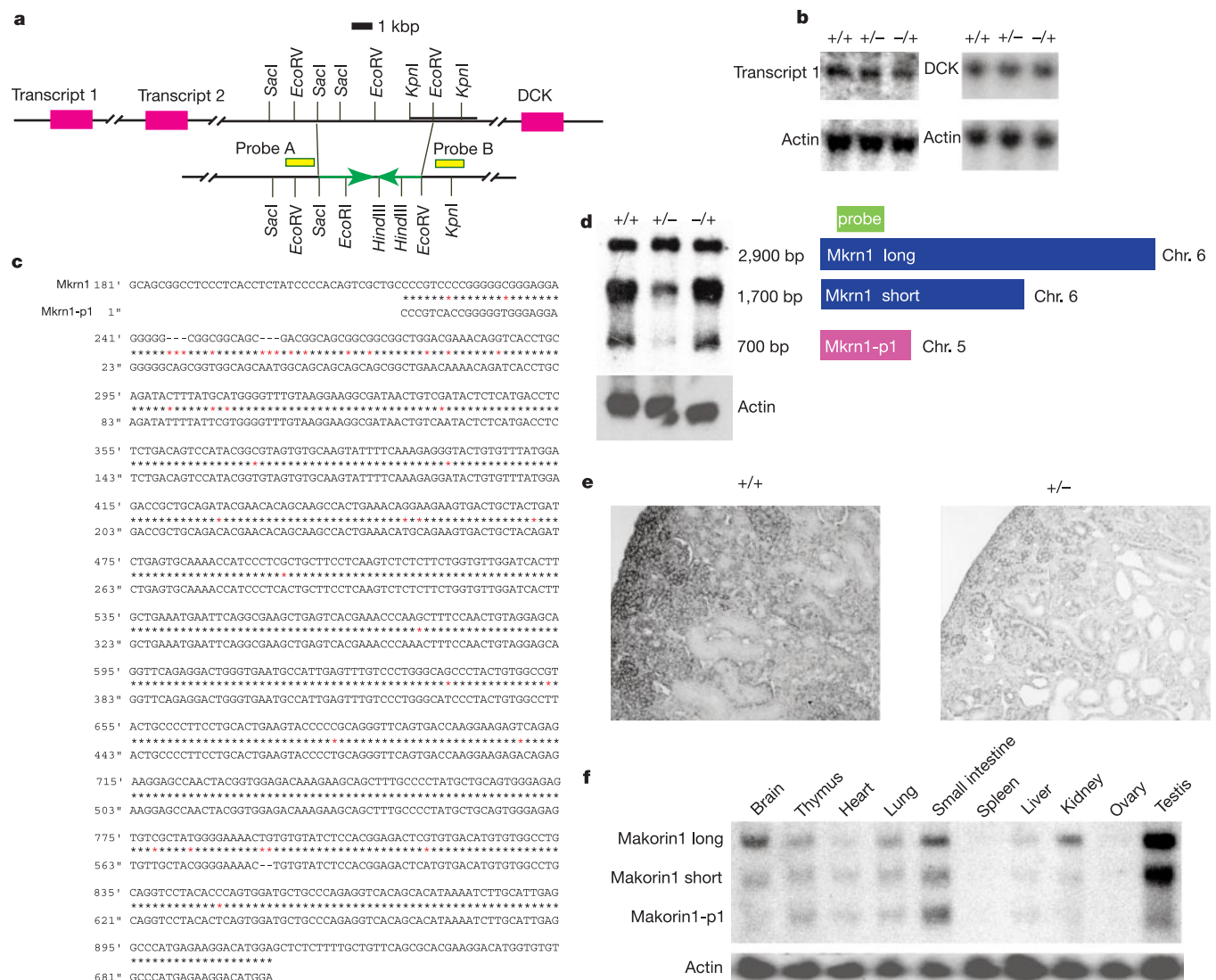


Figure 2 Identification of the mutated gene and expression profile. **a**, Diagram of the genomic organization of the transgene insertion locus, illustrating the transgene (green arrows), relative orientation of transcripts from three candidate genes (pink), and the probes used for screening of BAC library (yellow). **b**, Northern blot analysis of transcript 1 and DCK in heterozygous mice where the transgene was inherited from the father (+/-, affected) or mother (-/+, unaffected) revealed no obvious genotype differences. **c**, Sequence comparison of *Makorin1* (top) and *Makorin1-p1* (bottom), a pseudogene for

Makorin1. This was confirmed by the fragmented structure and accumulation of mutations (red asterisks) in *Makorin1-p1*. **d**, Northern blot analysis of *Makorin1* and *Makorin1-p1*, with a probe that can detect all transcripts. *Makorin1* gives rise to the two transcripts by alternative splicing, a long and a short form. **e**, Expression pattern of *Makorin1* in a wild-type (+/+) and a mutant (+/-) PO kidney. **f**, Northern blot analysis of several adult tissues.

Table 1 **Preferential expression of *Makorin1-p1* from the paternal allele**

♀	♂	MSM (%)	C3H (%)	n
MSM	C3H	27 (14%)	165 (86%)	192
C3H	MSM	161 (84%)	31 (16%)	192
♀	♂	<i>Spretus</i>	FVB (%)	n
<i>Spretus</i>	FVB	19 (10%)	173 (90%)	192
FVB	<i>Spretus</i>	174 (91%)	18 (10%)	192

Laboratory mouse strains C3H or FVB were mated to *M. spretus* or *M. molossinus* (MSM). mRNA of *Makorin1-p1* was amplified by RT-PCR, and 192 independent clones per embryo per genotype (in five genotype pairs) were isolated. Based on the restriction enzyme polymorphism, we determined whether each clone was derived from the paternal or maternal chromosome, and the number as well as percentage is displayed.

form is the main transcript of *Makorin1*. As *Makorin1-p1* RNA is known to be polyadenylated, we performed northern blot analysis with poly-A enriched mRNA extracted from E18 embryos using a 400-bp DNA fragment that can detect all transcripts (two transcripts of *Makorin1* and *Makorin1-p1*). Surprisingly, the short form of *Makorin1* as well as *Makorin1-p1* were significantly reduced in a heterozygous embryo derived from the mating between a mutant male and a wild-type female (+/-) compared to a wild-type embryo (Fig. 2d). In contrast, there was no significant reduction in these transcripts in a heterozygote derived from the opposite mating between a wild-type male and mutant female (-/+, Fig. 2d). Similar reduced transcript levels in mutants were found in embryonic (E 15.5) to early postnatal stages (P7). After maturation (over 4 weeks), this reduction of transcripts became less obvious, suggesting that epigenetic modification might be gradually erased (data not

shown). We performed methylation scanning of the genomic regions of *Makorin1* and *Makorin1-p1* by Southern hybridization coupled with cleavage of methylation sensitive restriction enzymes. The data suggested there was no obvious differential sites of methylation surrounding these regions (data not shown).

To determine whether the expression of *Makorin1-p1* is imprinted, we directly examined the paternal transcript and the maternal transcript of *Makorin1* and *Makorin1-p1*. There were no polymorphisms of *Makorin1-p1* in laboratory mice. Therefore, we mated laboratory mice (FVB) with wild mice (*Mus spretus* or *Mus molossinus*), and distinguished the two alleles by an enzyme polymorphism of polymerase chain reaction (PCR) products. As expected, the paternal allele of *Makorin1-p1* was preferentially expressed in either mating (Table 1). In contrast, no selective expression of the short form or the long form of *Makorin1* was observed (data not shown). These experiments were repeated five times from independent embryos, and were highly reproducible.

We further focused on expression in the kidney. In the wild-type kidney, renal vesicles and ureteric bud epithelium in the superficial cortical region were highly positive for *Makorin 1* expression, whereas inner differentiated and medullary regions were weakly positive or negative (Fig. 2e). Although signals were detected in the mutant, the staining pattern was weak and irregular in the affected tissue. Those data are roughly consistent with the phenotypes observed in the mutant. The expression profile of *Makorin1* was analysed by northern blot (Fig. 2f). While *Makorin1* was broadly expressed in various tissues, the expression level was significantly lower compared to the embryo, and high level of expression was only observed in the testis. Compared to the embryo, reduction of

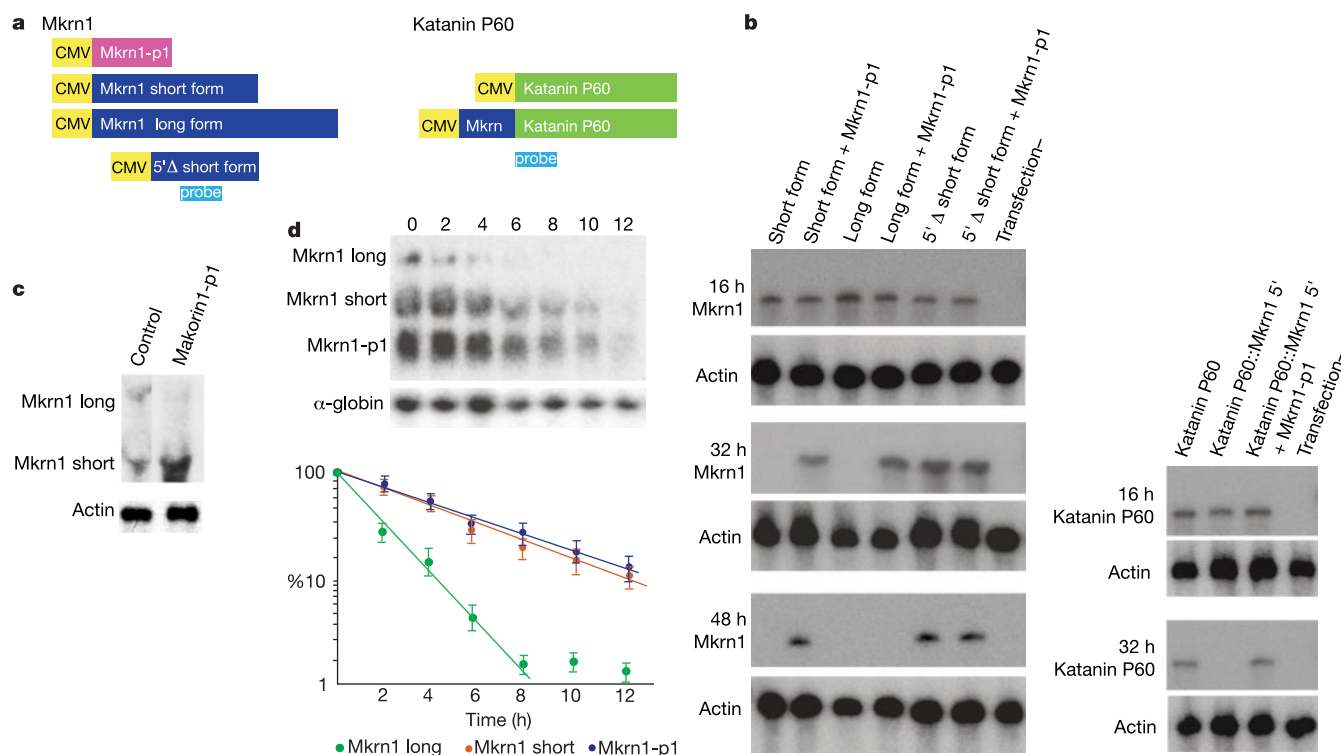


Figure 3 Demonstration of stabilization of *Makorin1* mRNA by *Makorin1-p1* mRNA in trans. **a**, Schematic presentation of *Makorin1-p1* and *Makorin1* long- and short-form CMV expression constructs, and the position of the protection probes. To demonstrate whether a *cis* regulatory element is present in the 5' region of *Makorin1*, we made a construct in which the 5' region of *Makorin1* was deleted (5'Δ), or conjugated to *Katanin P60*. **b**, **c**, Protection from degradation of *Makorin1* by *Makorin1-p1*. Plasmid combinations at top were transfected into NIH3T3 cells. The expressed mRNA was measured by RNA protection assay at 16 h, 32 h or 48 h after transfection. Actin was used

for control. **c**, The protective effect of *Makorin1-p1* on endogenous *Makorin1* was examined in P19 cells. **d**, The half-life of endogenous *Makorin1*s in P19 cells. After the start of incubation with actinomycin D, mRNA was extracted every 2 h and subjected to northern blot. Each time point was normalized to time 0, and plotted as a semilogarithmic graph for the decay of the long form of *Makorin1* (green circles), the short form of *Makorin1* (red circles) and *Makorin1-p1* (blue circles). Columns represent the mean of triplicate assays $\pm$ s.e.

expression of the short form of *Makorin1* and *Makorin1-p1* was especially notable.

We examined the mechanism underlying the ability of *Makorin1-p1* to regulate the stability of *Makorin1*. We made expression plasmids where the short form of *Makorin1*, the long form of *Makorin1*, or *Makorin1-p1* were driven by the cytomegalovirus (CMV) promoter (Fig. 3a). The short form or the long form of *Makorin1* was transiently expressed with or without *Makorin1-p1* in NIH3T3 in which endogenous *Makorin1* was not detectable by northern blot. mRNAs transcribed from the expression vectors were measured by an RNA protection assay. Although mRNAs transcribed from the short form or the long form of *Makorin1* provided robust signals at 16 h after transfection, these signals completely disappeared after 32 h, suggesting that the mRNAs of *Makorin1* were degraded by this time (Fig. 3b left). In contrast, when the short or long forms of *Makorin1* were co-transfected with *Makorin1-p1*, similar amounts of mRNA were preserved at 32 h as well as 16 h, suggesting that *Makorin1-p1* effectively protected both forms of *Makorin1* from degradation. We further extended this analysis to 48 h in culture, and examined the amount of RNAs. Curiously, after cotransfection with *Makorin1-p1*, the mRNA of the short form was present at that time, while the long form was not, suggesting that the short form was more efficiently protected by *Makorin1-p1*.

If *Makorin1-p1* regulated the stability of *Makorin1* mRNA in *trans*, we reasoned that the 5' end of *Makorin1* may be required for this regulation, because *Makorin1-p1* consists of the 5' 700 bp of *Makorin1*. Therefore, we made an expression vector carrying a truncated form of *Makorin1*, which lacked 700 bp of 5' region. This truncated *Makorin1* transcript was extremely stable in the NIH3T3 regardless of the presence or absence of *Makorin1-p1* (Fig. 3b left), suggesting that the 5' region of *Makorin1* contains a *cis*-acting element involved in the regulation of stability of mRNA. This was confirmed using a fusion mRNA in which 800 bp of *Makorin1* was linked to *Katanin P60* (Fig. 3a). The transcript from *Katanin P60* was stable at 16 h and 32 h. In contrast, the *Makorin/Katanin* fusion transcript disappeared at 32 h. Destabilization of the *Makorin/Katanin* fusion mRNA by the *Makorin1* 5' region was completely rescued by the expression of *Makorin1-p1* (Fig. 3b right). Next, we used the P19 cell line, which expresses

Makorin1 mRNA, to examine the effect of overexpression of *Makorin1-p1* on endogenous *Makorin1*. Both forms of *Makorin1* were expressed at low levels in the P19 cells. However, overexpression of *Makorin1-p1* significantly stabilized the short form of *Makorin1* (Fig. 3c). These data indicate that a *cis*-acting element that facilitates degradation of mRNA is present in the 5' region of *Makorin1*, and *Makorin1-p1* expression protects *Makorin1* from degradation in *trans*. In addition, protection of the short form of *Makorin1* appears to be more effective, consistent with observations in mutant mice and the RNA protection study (Fig. 2d, Fig. 3b).

We further investigated whether *Makorin1-p1* expression stabilized mRNA of the short form of *Makorin1*, or mRNA of both forms, under physiological conditions, by measuring the half-life of endogenous *Makorin1* mRNAs. P19 cells were treated with actinomycin D to inhibit nascent RNA transcription, and each mRNA half-life was measured at 2-h intervals. All *Makorin1* mRNAs were robustly expressed in the P19 cells (Fig. 3d). The stability of the endogenous short form of *Makorin1* and *Makorin1-p1* were tightly correlated (half-life, 4 h), whereas the long form of *Makorin1* decayed much faster (half-life, 1.8 h) regardless of the presence of *Makorin1-p1*. These data support the interpretation that *Makorin1-p1* selectively stabilizes the short form of *Makorin1*.

It was necessary to determine by transgene rescue whether the phenotypes of the mutant were due to lowered expression of the *Makorin1* gene resulting from the lowered expression of the unlinked pseudogene *Makorin1-p1* associated with insertion/deletion. Therefore, we first tested whether transgenic restoration of expression of *Makorin1-p1* could lead to increased expression of *Makorin1* and amelioration of the phenotypes. We also tested complementation of the phenotypes through direct overexpression of *Makorin1* from a transgene. We injected *Makorin1-p1* or *Makorin1* cDNA under the control of the ubiquitously expressed CMV/chicken beta-actin (CAG) promoter⁷ into fertilized eggs derived from the matings between heterozygous mutant males and wild-type females. After injection of each construct into approximately 400 fertilized eggs, we obtained 23 mice (21 mice) which were carrying transgenes of *Makorin1-p1* (*Makorin1*). Fourteen out of 23 mice (16 out of 21 mice) were heterozygous insertion/deletion mutants that also carried transgenes of *Makorin1-p1*

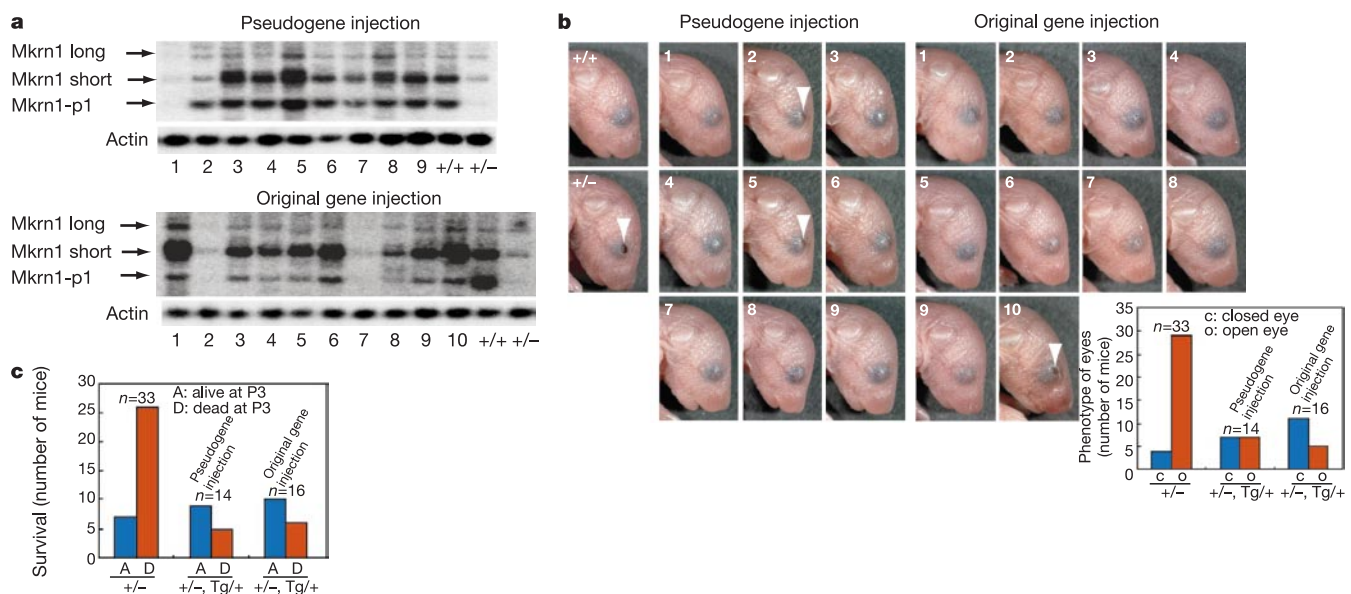


Figure 4 Rescue of the mutant phenotypes by exogenous *Makorin1-p1* or *Makorin1* cDNA transgene expression. **a**, Detection of the transgene expression by northern blot analysis using RNAs extracted from the kidneys of transgenic mice that survived for three days. Various degrees of expression from the transgene, as well as rescue of *Makorin1*

expression by the *Makorin1-p1* transgene, were observed. **b**, Rescue of the eye-closure defect in the mutants carrying the transgene. Each case is presented on the left, and summarized as a bar graph on the right. **c**, Survival three days after birth in mutant mice with and without the transgene.

(*Makorin1*). Although all wild-type mice carrying each transgene were grossly normal, a relatively low frequency of wild-type mice were found carrying transgenes, suggesting that overexpression of *Makorin1-p1* or *Makorin1* might affect normal development.

The expression of transgenes in the kidney of the insertion/deletion mutants that survived to three days after birth was detected by northern blot (Fig. 4a). While variable levels of expression from the transgenes were observed, the majority of either the *Makorin1* or *Makorin1-p1* transgenic mice expressed levels of the appropriate transgene that were close to the wild-type control. In addition, levels of the short form of *Makorin1* were restored by expression of the *Makorin1-p1* transgene, demonstrating *in vivo* what we had previously shown in cells in culture (Fig. 3).

To determine if there was rescue of the insertion/deletion phenotypes by either the *Makorin1-p1* or *Makorin1* transgenes, we examined the skin defect on the eye and the survival three days after birth. The eyelid skin closure defect of the insertion/deletion mutants carrying either transgene were ameliorated or normalized (Fig. 4b). Similarly, survival was substantially improved in the transgenic mice (Fig. 4c), although all transgenic mice did not survive to three days. The variability in expression of the transgenes suggests that poor expression of some of the transgenic mice may not be effective for rescue of lethality. These data strongly suggest that the phenotypes observed in the insertion/deletion mutant can be attributed to the lowered expression of *Makorin1* associated with the lowered expression of *Makorin1-p1* linked to the transgene insertion/deletion.

With these studies, we have uncovered a mechanism of gene regulation of a coding gene, *Makorin1*, by a transcribed non-coding pseudogene *Makorin1-p1*. *Makorin1-p1* must function as an RNA, as it cannot code for a protein. Protection from mRNA decay of *Makorin1* by *Makorin1-p1* was easily reproduced by expression constructs in several cell lines and in transgenic mice, suggesting that this type of regulation may be a general phenomenon. Although the precise enzymatic mechanism is unknown, one possibility is that a *trans*-acting RNA-destabilizing factor—which is able to bind to the 5' UTR of *Makorin1* and *Makorin1-p1* transcripts—is in limiting amounts or temporally and spatially regulated. Expression of the *Makorin1-p1* transcript may titrate out this factor, thus increasing wild-type *Makorin1* mRNA stability in a developmentally regulated and tissue-specific manner.

In general, the mammalian genome contains a large number of pseudogenes (20,000 in the human³) compared to other organisms. Although functions of pseudogenes have not been characterized, some proportion of them are transcribed⁸. Other expressing pseudogenes may have important functions, like *Makorin1-p1*. Pseudogenes are functionally less constrained, and have accumulated more mutations than translated genes. If they have some functions in gene regulation, this property would allow more rapid functional diversification than protein-coding genes. In addition, some genetic phenomena that exhibit incomplete penetrance might be attributed to 'mutation' or 'variation' of pseudogenes. □

Methods

Generation of a mutant mouse

Two constructs were injected to produce a transgenic mouse. First, we cloned a blunt-ended *NotI*-*Sall* fragment of *Drosophila Sxl* cDNA into a blunt-ended *NotI* site of pTRE2 (Clontech). Second, a *XhoI*-*SapI* fragment (4.75 kbp) from pTRE2 carrying *Sxl* and a *XhoI*-*HindIII* fragment (2.25 kbp) derived from pTet-Off (Clontech) were excised. Both constructs were injected into the pronucleus of fertilized eggs. The transgenic mice were bred with FVB to maintain an inbred state.

Breeding and genetic study

Heterozygotes were mated with wild-type FVB to maintain an inbred state. Genotyping was carried out by PCR. We used primers of TAGTTGTGTCGCTGGTAGAGATG and CCGAAATGCACATAATGAGTAATA for detection of the mutant allele, and primers of CGCGACTCTCCCTTTGTGGAC and TGGVVTGTTCTCTGCCTGTTACAC for detection of the wild-type allele. To demonstrate imprinted expression of *Makorin1-p1*, we amplified mRNA of *Makorin1-p1* by PCR with reverse transcription (RT-PCR) using a primer pair of GGCGGCGCTGGACGAACA and TGGGCAGCATCCACTGGGTGTAGG,

and cloned PCR products into Blue Script. We distinguished paternal or maternal alleles by the polymorphism restriction enzyme, *HindIII* and *BglII* (Toyobo). To avoid bias, we analysed 192 independent clones as one experimental procedure, and repeated this analysis five times from independent embryos.

Animal studies

All animal work was carried out at the animal facility of Saitama Medical School and the National Human Genome Research Institute under approved animal protocols and in accordance with institutional guidelines.

Northern blot and RNA protection assay

For northern blotting, poly-A enriched RNA was prepared from tissues or transfected cells by the same protocol (see above), run on a denaturing formaldehyde/0.8% agarose gel, and blotted on Hybond N+ (Amersham) according to the manufacturer's protocol. Probes were labelled with ³²P using Prime-It, a random primer labelling kit (Stratagene). To assay protection of mRNA by the pseudogene RNA, we made expression constructs by cloning both the full-length short form (EcoRI-*NotI* fragment) and long form of *Makorin1* (EcoRI-*NotI* fragment), as well as a truncated fragment (*HindIII*-*NotI* fragment) of *Makorin1*, into pCDNA3.1/His (Invitrogen). We also cloned either a full-length *Katanin P60* cDNA (EcoRI-*NotI* fragment) or the 5' end of *Makorin1* (EcoRI-*HindIII* fragment) fused to the full-length *Katanin P60* (EcoRI-*NotI* fragment) cDNA individually into pCDNA3.1/His. To generate an RNase protection construct, a *SacI*-*BglII* fragment of *Makorin1* or a *Sall*-*XhoI* fragment of *Katanin P60* was cloned into BlueScript. Transfection and RNA extraction was performed as described above. RNase protection was carried out according to standard protocols. To measure the half-life of *Makorin1*s, we cultured P19 cells in the presence of 10 µg ml⁻¹ of actinomycin D. RNA was extracted at 2-h intervals. Total RNA was subjected to northern blotting. Intensity of signals was measured by a BAS2000 instrument (Fuji). Each set of procedures was performed in triplicate.

Generation of transgenic mice using *Makorin1-p1* and *Makorin1* cDNA

Transgenic lines of mice for rescue experiments were generated by using the pCAGGS vector (provided by J.-i. Miyazaki). The *Makorin1-p1* pseudogene or the original *Makorin1* protein-coding cDNA was inserted into the *EcoRI* site of the vector. Inserted plasmids were partially digested with *Sall* and *PstI* (*HindIII*), and then electrophoresed. Each transgene was purified from agarose gel. *In vitro* fertilization was performed with sperms from mutant male mice and eggs from BDF1 female mice to obtain sufficient numbers of fertilized eggs. Transgenic mice were produced by microinjection of the transgene into fertilized eggs at the Laboratory Animal Resource Center, University of Tsukuba. The pCAGGS-GFP plasmid (a gift from M. Okabe) was co-injected with the transgene as a marker in transgenic mice. Surviving oocytes were transferred to the oviducts of pseudopregnant foster mothers. Finally, we injected each construct into approximately 400 eggs. The presence of the transgene was determined by PCR using primers TATTGTGCTGCTCATCATTTTGGC-TTGCTCTCTCTGTCTCTCTCT and TATTGTGCTGCTCATCATTTTGGC-CCCATAGCGACACTCTCCACGT for the *Makorin1-p1* transgene and the *Makorin1* transgene, respectively. Skin over the eyes of the mice was examined at birth. Survival rate was determined at three days after birth. After determination of survival rate, transgenic mice were killed to isolate kidneys for northern blot, western blot and histological examination.

Received 31 December 2002; accepted 17 February 2003; doi:10.1038/nature01535.

1. Vanin, E. F. Processed pseudogene: Characteristics and evolution. *Annu. Rev. Genet.* **19**, 253–272 (1985).
2. Goncalves, L., Duret, L. & Mouchiroud, D. Nature and structure of human genes that generate retropseudogenes. *Genome Res.* **10**, 672–678 (2000).
3. Harrison, P. M. *et al.* Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res.* **12**, 272–280 (2002).
4. Nagai, Y. *et al.* Role of interleukin-6 in uncoupling of bone in vitro in a human squamous carcinoma coproducing parathyroid hormone-related peptide and interleukin-6. *J. Bone Miner. Res.* **13**, 664–671 (1998).
5. Gray, T. A. *et al.* The ancient source of a distinct gene family encoding proteins featuring RING and C (3) H zinc-finger motifs with abundant expression in developing brain and nervous system. *Genomics* **25**, 14430–14435 (2000).
6. Greally, J. M. *et al.* Conserved characteristics of heterochromatin-forming DNA at the 15q11-q13 imprinting center. *Proc. Natl. Acad. Sci. USA* **25**, 14430–14435 (1999).
7. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–200 (1991).
8. Mighell, A. J., Smith, N. R., Robinson, P. A. & Markham, A. F. Vertebrate pseudogenes. *FEBS Lett.* **468**, 109–114 (2000).

Acknowledgements We thank Y. Sugimoto, T. Matsukawa, T. Harigaya, N. Deguchi, Y. Obata, K. Moriawaki and M. Muramatsu for support; T. Itoh, T. Matsumoto, S. Sasaki, M. Ishida and Y. Yamauchi for technical support; P. Schedl for providing *Sxl* cDNA; T. Shiroishi for providing *M. molossinus*; J.-i. Hata for comments on pathology; G. G. Germino for suggestions and encouragement; S. Okumura, K. Hagiwara, N. Tominaga and M. Suzuki for mouse breeding; and S. Kunita, N. Kajiwara, K. Furuya, M. Hirose, H. Sanno and Y. Suzuki for generating transgenic mice. This work was supported by PRESTO, Japan Science and Technology Corporation (S.H.), PROBRAIN (S.T.), and Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, and Technology (A.Y.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H. (shinji@saitama-med.ac.jp).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Seeking out the élite

Scientists seeking a job at an institution that boasts a strong impact factor could research their prospects by following citation trails back to their source. Perhaps not surprisingly, such trails lead to prominent institutions in well-established scientific regions, according to the company ISI in Philadelphia, which tracks citations.

Earlier this year, ISI's publication *Science Watch* performed an evaluation to see which institutions dominated citation statistics in molecular biology and genetics for the period 1992–2002. The analysis generated two lists: efficiency and power. The former was based on the percentage of all papers published that received a high number of citations; the power list looked at the raw number of cited papers per institution. Cold Spring Harbor Laboratory in New York won the efficiency contest, whereas Harvard University powered through to head the second category.

ISI's efficiency rankings show where the hot molecular-biology regions are in the world. Twelve of the top 15 are clustered on the east or west coast of the United States. Only two non-US institutes made the top 15 — the European Molecular Biology Laboratory in Heidelberg, Germany, and the Imperial Cancer Research Fund (now called Cancer Research UK) in London. The 'power' list fills in the gaps, with a few more institutions that are located near those with the high efficiency ratings. It also adds other institutions to the mix, such as Duke University in Durham, North Carolina, the Pasteur Institute in Paris and Kyoto University in Japan, that have historically been strong in molecular biology but which aren't part of the major US coastal hubs.

Of course, individuals, not institutions, earn citations. Although places that have scored high on either list would be justifiably proud in trumpeting their rankings, young scientists would be wise to check their prospective mentors' citation records.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS AND STUDENTS

Taking control of
your destiny

p98

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Getting organized

Postdoctoral associations on both sides of the Atlantic are mobilizing to tackle long-standing problems and smooth the path through this transitional phase in a scientist's career. Sally Goodman and Karen Kreeger report.

More than 200 past and present Marie Curie fellows discuss progress in Paris.

Salary, mobility, benefits: postdocs in Europe and the United States share many of the same concerns. And postdoc organizations on both continents are addressing them in an increasingly organized manner. Their recent meetings have attracted high turnouts, although only time will tell whether their plans will lead to progress.

In Europe, the Marie Curie Fellowship programme — originally set up to encourage mobility — is being reshaped in consultation with past and present fellows. A meeting in Paris this March gave current fellows a chance to hear how the European Commission (EC) was responding to feedback on their mobility experiences. And a Brussels meeting attended by an EC official and the Marie Curie Fellowship Association (MCFA) allowed past fellows to discuss what further changes they would like to see.

In California this March, the US National Postdoctoral Association (NPA) held its inaugural meeting, aiming to influence programmes in the

United States. But it struggled with the changing conception of a postdoc — from the old notion of someone in the midst of a training grant, to the more amorphous but accurate picture of anyone who is seeking a permanent career after completing a PhD.

EUROPEAN REUNION

The plight of Marco Vecoli, an Italian postdoctoral researcher in palaeontology at the University of Brest in France illustrates both the promise and problems with the old Marie Curie Fellowships, which only funded researchers to work outside their own country. Vecoli is happy to have had consecutive Marie Curie Fellowships in Germany and in France, but would like to return to a job in Italy after four years away.

Vecoli was one of 200 past and present Marie Curie fellows from 22 countries who attended a feedback meeting at the Curie Institute in Paris on 12–14 March. The session was organized with the aim of making the grants more attractive — especially to young researchers concerned about finding work in their home country after taking a fellowship abroad.

Georges Bingen, head of the Marie Curie Fellowships unit at the EC, acknowledges that repatriation has been a problem in the past. “We have to encourage young researchers to leave their home country knowing that they will be able to return,” he says. But he believes that the new-look Marie Curie programme for human resources and mobility, part of the EC's Sixth Framework Programme for research which runs until 2006, addresses this issue — basically, by spending more money on fewer grants.

With a E1.58-billion (US\$1.7-billion) budget for mobility-related activities — an increase of more than 50% over the previous programme, which awarded 12,000 fellowships — the EC hopes that it will be able to incite some 8,500 researchers to pack their cases, but for longer periods away, and with more guarantees of a happy return. All postdoctoral fellows can apply for reintegration grants of up to E40,000, aimed at supporting research in the institution that appoints

Marco Vecoli: happy to have spent four years abroad, but now hoping to return to his native Italy.



them on their return from a stint abroad.

Marie Curie fellows have also had problems in going to the country of their fellowship. Difficulties securing housing and health cover can be a nightmare in administration-heavy countries such as France, particularly for researchers coming from outside the European Union. The EC has taken heed of these concerns in the new programme by setting up a web-based mobility portal and national mobility centres — one-stop shops aimed at giving country-specific advice to researchers before they leave home.

Magda Lola, a member of the MCFA's board, says that fellows are generally positive about the changes made to the programme. "With no age limit and with fellowships open to countries outside of Europe, it is a much more flexible scheme," she says.

But some researchers fear that broadening the programme's range may slow down the application process for fellowships, already considered by some to be too long. "Science moves fast. We can't always wait to have the results of the fellowship before moving," says German postdoc Claus Fütterer, who has a Marie Curie Fellowship in biophysics at the Curie Institute.

Others question the requirement to secure a post before they can benefit from a reintegration grant. But Bingen insists that responsibility for mobility must be shared. "It's not the commission's job to pay for life-long salaries for researchers," he says. "The responsibility to provide jobs must be at the national level. But we can help to make them attractive."

The EC has put up money for mobility. But with only 5% of the finance for postdoctoral research in Europe coming from commission funds, the challenge is now for governments to make a more obvious commitment to the cause.

UNITED STATES

In the United States, the biggest problem isn't funding of national research, it's the lack of consistency on benefits such as stipends and insurance. Benefits tend to vary from institution to institution — sometimes even when they are funded by the same agency.

It was to address these and other issues that the NPA held its inaugural meeting in Berkeley in March. "We must develop compelling strategies for implementing fair and equitable policies that are good for everyone," says Orfeu Buxton, a postdoctoral fellow in neuroendocrinology at the University of Chicago and a founding member of the NPA.

At the meeting, the association began work on a



white paper of recommended postdoc practices. The association is soliciting input from members through its website, as well as from local postdoctoral associations. The group will eventually present its findings to the National Institutes of Health — one of the biggest funders of postdocs in the United States, and an organization that others look to for precedents.

Another important issue is the changing definition of a postdoc. The old criterion of a fixed-term training period no longer applies, says Xenia Morin, a Keck postdoctoral fellow and lecturer in biology and chemistry at Bryn Mawr College in Pennsylvania, who gave a talk on the subject at the NPA meeting. Morin, a teaching postdoc (see *Naturejobs* 5; 28 February 2002), says she is treated like a colleague. She favours a broad definition, encompassing anyone in transition from graduate studies to a permanent position.

No matter what route postdocs are taking, she finds that the line between training and employment is becoming increasingly blurred. Organizations such as the MCFA have shown that European postdocs can make the transition easier if they mobilize and persist. The NPA will soon learn whether US postdocs can effect similar changes when they band together. ■

Sally Goodman is a freelance writer based in Paris; Karen Kreeger is a freelance science writer based in Philadelphia.

Xenia Morin favours a broad definition of postdoc that could include anyone in transition from graduate studies to a permanent position.

Happy returns

The Marie Curie fellowships, as envisioned under the European Commission's Fifth Framework Programme, at times seemed to encourage movement for the sake of movement. The Sixth Framework, which came into effect late last year, aims to correct that, says Raffaele Liberali, who directs Marie Curie activities at the

European Commission.

One of the biggest changes is that mobility-encouraging grants are no longer emphasized just for younger researchers in the Marie Curie programme. Opportunities to travel are now extended to researchers of all ages and levels of experience.

Also, the European Commission

is now more conscious of brain drain — especially from non-member countries.

"The return element for researchers coming from non-European countries is particularly crucial," says Liberali, stressing the need for scientists to go back and aid the development of research in their home countries. S.G.

Web links

US National Postdoctoral Association

♦ www.nationalpostdoc.org

Marie Curie Fellowship Association

♦ www.mariecurie.org

Marie Curie Actions

♦ europa.eu.int/comm/research/fp6/mariecurie-actions/home_en.html

mariecurie-actions/home_en.html